Gene regulatory network of melanocyte development

Lapedriza, Alberto

Award date:
2016

Awarding institution:
University of Bath

Gene regulatory network of melanocyte development

Alberto Lapedriza

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

October 2015

Table of Contents

Acknowledgments

Firstly, I would like to thank my “biology” supervisor Prof Robert Kelsh for his support, enthusiasm and patience. I would also like to thank my “maths” supervisor Dr Hartmut Schwetlick for all his help and support with all the mathematical related aspects of the project.

Thank you everyone from Lab 0.76, past and present, who have helped and supported me, and for the best coffee breaks discussions. Special thanks to Marc Shedden for his expert fish care.

Finally a huge thank you to my friends and family, for putting up with me and for their continuous support in every single way. This work would have not been possible without you.

Abstract

Greenhill *et al.* (2011) developed a gene regulatory network of the main genes and interactions known to play a role in melanocyte biology, and generated a mathematical model to describe the behaviour of this complex network using semi quantitative data (ISH expression data). In this project we sought to collect expression data from four genes of the melanocyte GRN (*sox10*, *kit*, *mitfa* and *dct*) to develop a quantitative model that is able to describe the data more accurately. Moreover, we intended to identify more genes that are part of the melanocyte development process to be incorporated to the GRN.

We analysed microarray data that compared differentially expressed genes between *sox10* mutant and wild type embryos and validated five genes with a key role in melanocyte biology as downregulated in *mitfa* mutant embryos, which are downstream of *mitfa* in the GRN. We suggest that *kit* plays the role of factor Y in the Greenhill *et al.* (2011) GRN: Mitfa drives *kit* expression, and *kit* expression is transiently driven by Sox10 at early stages of development. As part of the feedback loop, *kit* seems to drive and maintain *mitfa* expression, however this needs to be validated. Finally we developed an experimental set up to obtain an estimate of gene expression per melanocyte from *sox10*, *kit*, *mitfa* and *dct*, using both qPCR and ISH cell count measurements. With this estimate we performed a parameter optimisation procedure, and found a set of parameters for the mathematical model that predicted the experimental data very accurately. The new model suggests that low expression values of *sox10* are sufficient to drive *mitfa* expression in high levels. It also predicts that high expression of *sox9b* is needed to achieve the high expression levels of *dct* seen in the data, although these predictions need to be experimentally tested.

This study represents the first attempt to obtain fine-scale gene expression data from melanocytes for the development of a quantitative mathematical model in zebrafish.

Abbreviations

µg	microgram
µl	microlitre
µM	micromolar
bp	base pair
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
DNA	Deoxyribonucleic Acid
dpf	days post fertilisation
FACS	Fluorescence Activated Cell Sorting
FDR	False Discovery Rate
GFP	Green Fluorescent Protein
GRN	Gene Regulatory Network
HCM	High Content Microscope
HM	Hybridisation Mix
hpf	hours post fertilisation
ISH	*in situ* hybridisation
mRNA	messenger Ribonucleic Acid
NC	neural crest
NCC	neural crest cell
NTC	No Template Control
ODE	Ordinary Differential Equation

PCR	Polymerase Chain Reaction
qPCR	Quantitative RT-PCR
RNA	Ribonucleic Acid
TF	Transcription Factor
WT	Wild Type

Chapter 1:

Introduction

Introduction to development

Multicellular organisms do not originate fully formed, they arise from a series of progressive changes called development. The development of a multicellular organism starts with a single cell (the zygote) which divides to produce all the cells of the organism.

Developmental biology tries to answer essential questions in biology such as how different cell types organise themselves to form tissues and organs, how our cells know when to stop dividing and growing, or how to generate reproductive cells with all the information to make a new organism. However, the most fundamental process that developmental biology is trying to understand (and the one we are focusing on in this project) is the question of cell differentiation. A single cell, the fertilised egg, gives rise to hundreds of different cell types—muscle cells, epidermal cells, neurons, lens cells, lymphocytes, blood cells, fat cells, and so on. We need to understand how all these different types of cells are produced from the same set of genetic instructions (since each cell of the body, with very few exceptions, contains the same set of genes) (Gilbert, 2000).

In vertebrates the first step in development is the union of male and female gametes to produce the zygote, a process called fertilisation. Following fertilisation, the zygote starts dividing into a larger number of cells. This will result in a tightly packed mass of about 32 cells that is called morula, with each individual cell referred to as a blastomere. As the blastomeres continue to divide, they secrete a fluid into the centre of the morula. Eventually, a hollow ball of 500 to 2000 cells, the blastula or blastocyst in mammals, is formed. The pattern of cleavage division is influenced by the presence and location of yolk, which is abundant in the eggs of many vertebrates. In a mammalian blastocyst, each cell is in contact with a different set of neighbouring

cells, and these interactions will influence the developmental fate of each cell. This positional information will set up different patterns of development along three embryonic axes: anterior-posterior, dorsal-ventral, and proximal-distal. After the completion of cleavage, certain groups of cells invaginate from the surface of the blastula in a carefully orchestrated activity called gastrulation. By the end of gastrulation, the cells of the embryo have rearranged into three primary germ layers: ectoderm, mesoderm, and endoderm; with very different developmental fates. In general, cells from the ectoderm are destined to form the epidermis and neural tissue; the mesoderm gives rise to connective tissue, skeleton, muscle, and vascular elements; and the endoderm forms the lining of the gut and its derivatives. The events of gastrulation are one of the first fate choice events in the differentiation process (Gilbert, 2000).

More than half a century ago, Waddington proposed an epigenetic landscape to describe the cell differentiation process (Figure 1). In this figure a differentiating cell is represented by a ball, and it starts out in a valley at the back of the landscape. As the ball rolls forward and downward, the valley splits or bifurcates into two new valleys separated by a ridge. These new valleys represent alternative cell fates. External stimuli and/or internal genetic influences, determine which of the two valleys a particular cell chooses. The valleys continue to split, and eventually the cell ends up in one of many terminal sub-valleys, which represent terminally differentiated states. The cell is held permanently in its terminally differentiated state by high valley walls (Ferrell, 2012).

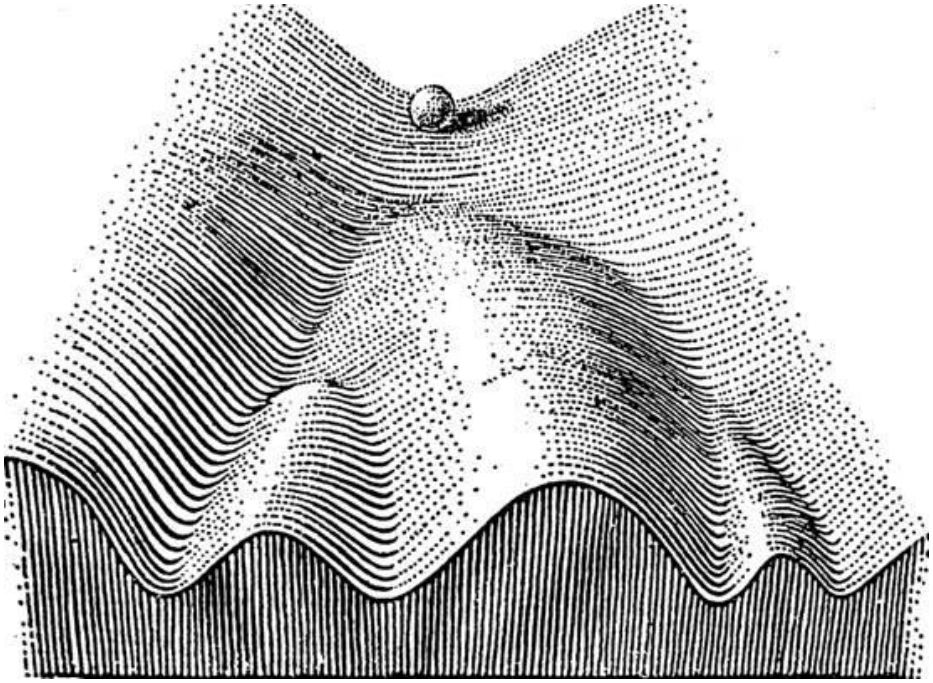

Figure 1. Waddington's epigenetic landscape. Differentiating cells are represented by a ball rolling through a surface. Valleys represent alternative differentiated cell states, and ridges keep cells from switching fates (Waddington, 2014).

However, there are some fundamental differences between Waddington's view of what happens to the landscape during differentiation, and what happens according to the current models based on experimentation. In Waddington's landscape, a cell begins at the bottom of a single well, and then as development proceeds this well successively splits into many more, representing the possible differentiation states of the cell. Therefore his model allows the reversibility of developmental events. However, the current models propose that differentiation mainly involves the disappearance of valleys from the landscape, not the appearance of new valleys. And this valley disappearance provides an irreversibility to the process of differentiation that was missing from Waddington's original view (Ferrell, 2012). Despite these differences his view is still attractive and useful for understanding development.

Cells contain many components, including genes, proteins, and metabolites. The cellular state at a particular time is determined by the abundance of its components at that time. Interactions among these components, such as the activation and repression of gene expression, causes changes in the cellular state, a phenomenon

that can be depicted as a trajectory in the Waddington's landscape, which represents differentiation (Furusawa and Kaneko, 2012).

Those interactions between genes, transcription factors and other components of signalling pathways form an intricate and network-like architecture known as a gene regulatory network (GRN). They are maps that state in detail the inputs and outputs into the elements regulating a developmental program. They also provide predictions on how the relations will evolve along the developmental program. The GRNs provide an explanation of the genetics interactions that drive differentiation (Levine and Davidson, 2005).

Neural crest development

The neural crest (NC) is a highly migratory multipotent cell population that forms in the early stages of embryonic development. Neural crest cells originate when neural tissue is formed in the ectodermal (the outer) layer of the developing embryo. The ectoderm is divided in three regions: the neural ectoderm or neural plate that will produce the central nervous system; the non-neural ectoderm, which will give rise to the epidermis; and the cells at the border of both regions that will form the neural crest cells. This neural tissue invaginates to form the neural tube, a process known as neurulation (Figure 2). During this process the neural plate border bends to form the neural fold and then the dorsal part of the neural tube (Bronner-Fraser, 1995, 2002; Gammill and Bronner-Fraser, 2003). This process might be slightly different depending on the organism: in zebrafish the process initiates with the columnarisation of an existing epithelium in the ectoderm to form the neural plate, which then forms a neural keel and solid tube; subsequently the lumen of the neural tube opens from ventral to dorsal (Lowery and Sive, 2004).

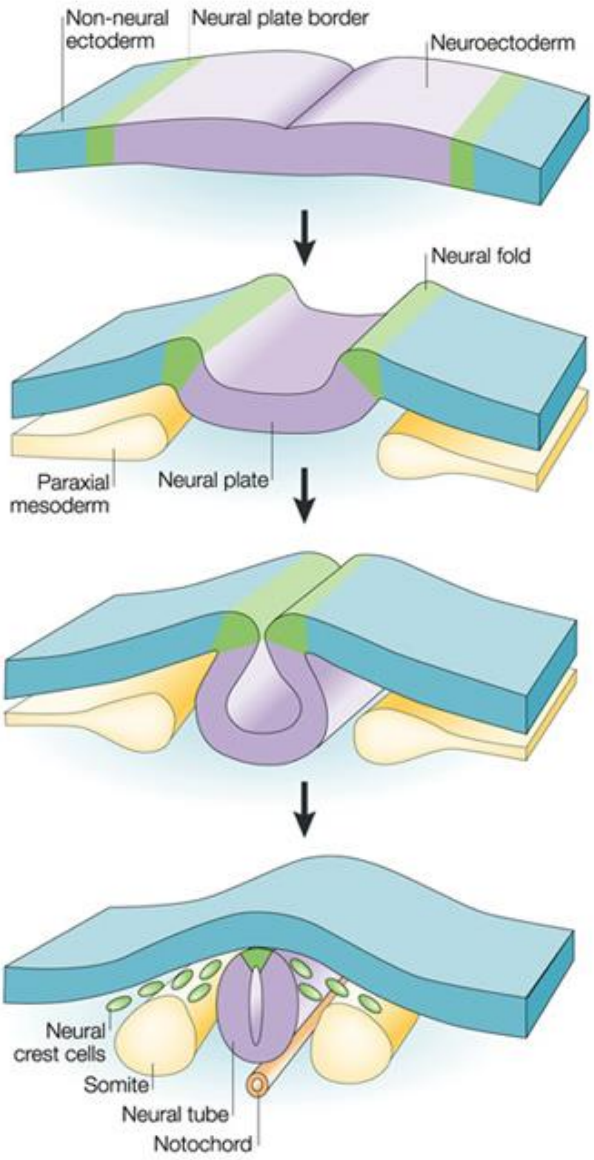

Figure 2. Neural crest formation. The figure shows the neurulation process, in which the neural plate folds to form the neural tube. Neural crest cells originate from the closing neural fold or from the dorsal part of the neural tube, depending on the organism (Gammill and Bronner-Fraser, 2003).

Specification of NC in the neuroectoderm is initiated in response to some factors, like BMPs, FGFs and WNTs which drive the expression of transcription factors involved in neural crest cells specification. BMPs have been shown to be sufficient to induce neural crest in birds (Liem *et al.*, 1997). However, additional work showed that neural crest formation requires BMP signalling only after the initial induction step, indicating that BMPs might have a maintenance role in the induction process, or that they signal

the emigration of neural crest cells from the neural tube (Sela-Donenfeld and Kalcheim, 1999). It now seems that the inducing signal from the non-neural ectoderm is a Wnt protein, and that Wnts are both necessary and sufficient for robust induction of neural crest in isolated neural tissue (García-Castro *et al.*, 2002). However, the role of Wnts and BMPs seems to be slightly different in zebrafish. It appears that in zebrafish a gradient of both BMP and Wnts is behind neural crest cell specification (Dorsky *et al.*, 1998). FGF signalling can also induce neural crest in neuralised ectoderm, although through a Wnt intermediary, as was proven in Xenopus (Gammill and Bronner-Fraser, 2003).

These factors initiate changes in cell adhesion and motility promoting their delamination from the neural tube in a process called epithelial-to-mesenchymal transition (EMT). NC cells will migrate through specific paths towards their destination (a process controlled by adhesion molecules and cues) while they start to differentiate into the appropriate cell type (Ruhrberg and Schwarz, 2010).

During and after migration, cells differentiate into a very wide range of derivatives that are grouped into two categories: ectomesenchymal, which include bone, cartilage, dentine and adipocytes (Billon *et al.*, 2007); and non-ectomesenchymal, including neurons, glia and pigment cells (Figure 3). Their multipotency (they give rise to many different cell types) and migratory ability (they are deployed at a wide variety of sites in the embryo) make them a very interesting and useful model to understand differentiation and development (Donoghue *et al.*, 2008).

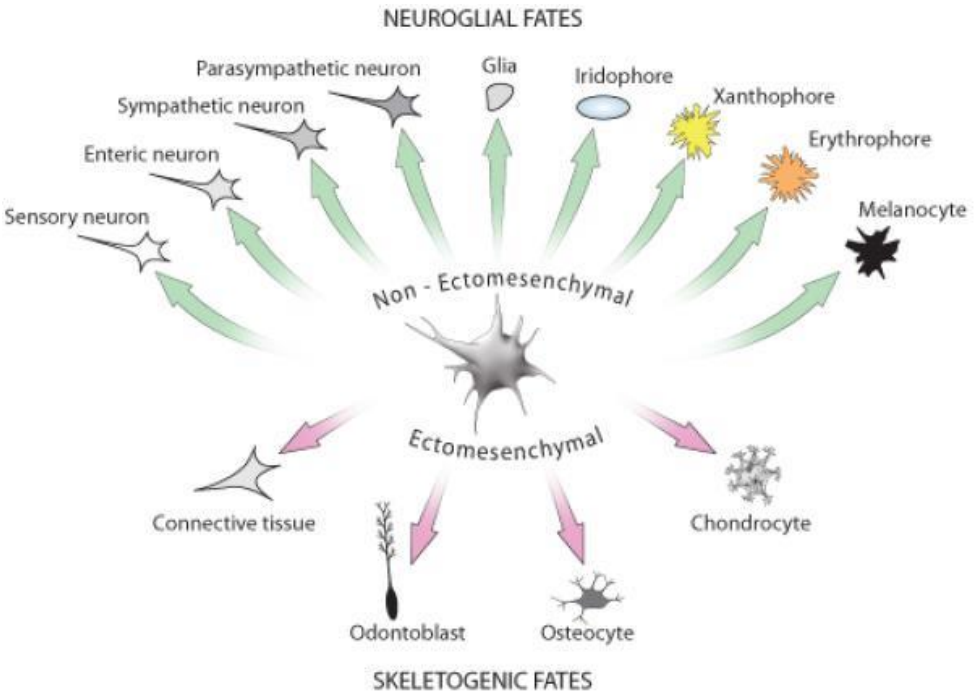

Figure 3. Neural crest derivatives. NC cells give rise to pigment cells, several neuron types, glia and skeletogenic cells (Donoghue *et al.* 2008)

After their delamination from the neural tube, NC cells disperse away and organise into different subgroups. Their migration direction is guided by cell to cell interactions, chemotactic signals and extracellular matrix. In the head NCC organise in different subpopulations that will migrate to specific locations to give rise to glia, neurons, cartilage or bones, among others. In the trunk two main migration pathways are observed (Figure 4): the dorsolateral pathway, mainly used by pigment cell precursors; and the ventromedial pathway primarily used by glial and neuronal precursors (Theveneau and Mayor, 2014).

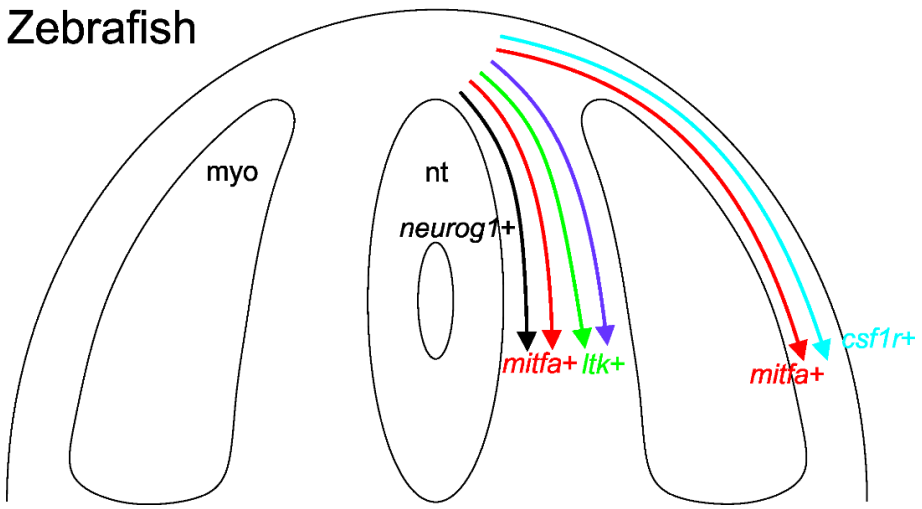

Figure 4. Migratory pathways of NCC derivatives. The schematic shows a transverse section of a zebrafish embryonic trunk and outlines the migratory pathways followed by NC derivatives with representative key marker genes. Red arrows represent *mitfa*-positive melanocyte precursors. Green and blue arrows represent iridophore and xanthophore precursors, respectively. Black arrows represent Neurogenin1 positive neural progenitors, which primarily generate both neuronal and glial components of the dorsal root ganglia. The purple arrows represent further glial progenitor types. dm, dermomyotome; myo, myoseptum; nt, neural tube (Lapedriza *et al.*, 2014).

Melanocytes as a model for development

The complexity of the genetic interactions that regulate the differentiation and specification of a neural crest derivative from their progenitor make these cells a very interesting and powerful model to understand cell differentiation and stem cell fate choice. As a multipotent stem cell differentiates, its GRN becomes reconfigured from a state where multiple fates can be chosen to a more stable state that indicates commitment to a specific cell fate. We still lack a deep understanding on how these changes in the cell gene regulatory network are produced (Levine and Davidson, 2005; Greenhill *et al.*, 2011)

In our lab, we are particularly interested in the genetic regulations which control the differentiation of the pigment cells from neural crest. Among all the pigment cells produced from the neural crest cells, melanocytes are one of the most useful and

interesting to study this fundamental aspect of development: they are large cells, well-characterised genetically and easy to identify (because of their colour) especially in zebrafish embryos which are transparent ; they are also affected in many human pigmentation diseases.

Zebrafish is a very suitable model to study developmental genetics. One of the main reasons for that is that the transparency of its embryos allows to observe how melanocytes develop better than in any other of the main genetic model organisms. The rapid development of melanocytes and pigment cells gives us a fast access to biological material in comparison to other organism. Furthermore, the evolutionary conservation of melanocytes in other organisms suggests the conservation of the genes and interactions responsible for melanocyte development (Sato *et al.*, 2001; Hallsson *et al.*, 2004). Lastly, we possess many zebrafish pigmentation mutants that have been and still are an invaluable tool to understand melanocyte development and development in general (Henion *et al.*, 1996; Kelsh *et al.*, 1996; Odenthal *et al.*, 1996; Gaiano *et al.*, 1996; Johnson *et al.*, 1995; Patton and Zon, 2001).

Neural crest-derived pigment cells (or chromatophores) are very diverse and conserved in evolution with fish, reptiles and amphibians having up to at least 6 pigment cell types (Mellgren and Johnson, 2002), including melanocytes (usually black or brown), iridophores (reflective silver, blue or gold), leucophores (white) and xanthophores/erythrophores (yellow/red). Mammals and birds have secondarily lost most of this diversity, retaining only melanocytes (and perhaps iridophores in birds; (Bagnara and Matsumoto, 2007)). Different pigment cells are characterised because they contain different pigment molecules (which they synthetise or obtain from the diet) usually stored in specialised pigment organelles. Pigment cells are usually associated with the skin, residing in the dermis and/or epidermis (Schartl *et al.*, 2015).

Melanocyte genetics is perhaps better characterised than that of any other cell-type (Lamoreux *et al.*, 2010). In contrast, other pigment cells have been less studied. Xanthophores are usually highly dendritic cells that appear yellow under white light. They contain granules of pteridine pigment, synthesised from guanosine triphosphate through complex and incompletely defined biochemical pathways

(Ziegler, 2003). Erythrophores are a closely-related cell-type and, although appearing red or orange, are sometimes difficult to distinguish from xanthophores as their appearance depends on the quantity and type of pteridine derivatives present within their granules (Goodrich *et al.*, 1941; Matsumoto, 1965). Both cell-types also may contain carotenoid granules: membrane bound organelles containing dietary carotenoids (Ichikawa *et al.*, 1998). Iridophores (appearing blue, silver or gold), and leucophores (white or cream) are non-dendritic cells that contain organelles termed reflective platelets, composed of crystalline guanine. The distinctive appearance of these two cell-types is primarily due to the organisation of the reflecting platelets; in leucophores they are found in all orientations, whereas in iridophores the spacing and orientation of the platelets is carefully orchestrated to generate the iridescence by a thin layer interference mechanism (Morrison, 1995). Cyanophores were identified in mandarin fish as blue pigment cells, owing to a so-far uncharacterised cyan biochrome within their granules (Goda and Fujii, 1995). Finally, irido-erithrophores are a recently discovered pigment cell that contains both reflecting platelets and vesicles containing an uncharacterised violet pigment (Goda *et al.*, 2011).

Melanocytes are characterised by the production of melanin, a high molecular weight polymer derived from tyrosine, within membrane-bound organelles known as melanosomes. In most species, melanosomes remain exclusively within the melanocytes, but often they can be moved around within the cell along microtubules, a process with a key role in the rapid colour changes so characteristic of fish and reptiles (e.g. chameleon). In contrast, in the skin of amniotes (birds and mammals), melanosomes are exported from melanocytes to adjacent keratinocytes; skin colour and tanning reactions are mostly due to this process of epidermal melanosome transfer (Lin and Fisher, 2007). Melanosome transfer is also prominent in the hair follicle where melanin is deposited in the developing hair (made from keratinocytes) as it grows. To facilitate this, mammalian melanocytes tend to be highly dendritic cells whose cell body rests on the basal lamina; their dendrites project into different layers of the epidermis where they are intimately associated with the keratinocytes of the skin (Kippenberger *et al.*, 1998).

Melanocytes produce two kinds of melanin: black to brown eumelanin and yellow to red pheomelanin, although the latter is thought to be restricted to mammals and birds (Ito and Wakamatsu, 2003). Eumelanin acts as a photoprotective anti-oxidant, whereas pheomelanin is phototoxic and a pro-oxidant (De Leeuw *et al.*, 2001). In mammals, natural melanin pigments consist of a mixture of both eumelanin and pheomelanin in varying proportions, with the overall colour reflecting the bias in ratios (Simon *et al.*, 2009).

Our understanding of the genetic regulation of melanocyte development has improved throughout the last decades as many new genes have been identified and studied. Moreover the various genetic screens in zebrafish have supplemented that understanding (Henion *et al.*, 1996; Odenthal *et al.*, 1996; Kelsh *et al.*, 1996; Gaiano *et al.*, 1996; Patton and Zon, 2001) and have finally allowed insight into the origin of and genetic mechanisms controlling other pigment cell types.

Pigment cells originate through progressive fate restriction from neural crest and they may share a partially restricted progenitor common to all pigment cell types, i.e. a chromatoblast. Large-scale mutagenesis screens in zebrafish embryos have identified numerous loci that affect all three pigment cell-types (Henion *et al.*, 1996; Kelsh *et al.*, 1996; Gaiano *et al.*, 1996). For example, *sox10* mutant displays a severe reduction in all pigment cell-types, which may imply that this gene has a broader role in the neural crest (Kelsh *et al.*, 2000). Clonal analysis studies of premigratory trunk neural crest cells in wild-type zebrafish embryos show that individual clones expressing these genes may contribute to more than one pigment cell-fate. Interestingly, clonal analysis in *sox10* mutants indicated that all pigment cell progenitors, but not neural progenitors, remained stuck in a premigratory position (Dutton *et al.*, 2001). The phenotype of the *mitfa* mutant in zebrafish combines absence of melanocytes with increased iridophores, which is interpreted as resulting from redirection of fate choice of multipotent precursors, although this remains to be formally tested (Lister *et al.*, 1999; Curran *et al.*, 2009). A recent study relied on co-visualisation of Mitf:GFP cells and different chromatophore lineage markers to suggest the existence of a bipotent melano-iridophore precursor (Curran *et al.*, 2010). These studies suggests that there are close links between different pigment

cells, however one question that remains unanswered is if there is just one common chromatoblast progenitor, or several bi-potent pigment cell progenitors.

Specification and fate commitment of pigment cell types from the NC forms an attractive model system for understanding the cellular and genetic basis of stem cell development. Furthermore, the importance of hair, skin and eye pigmentation in humans, and the numerous human congenital diseases result from disruption of pigment cell development make melanocyte an important cell to study. Further understanding of melanocyte development and the underlying gene regulatory networks will, therefore, not only help clarify the principles of stem cell regeneration and differentiation but will also confer significant medical insight.

Melanocyte biology

Melanoblasts, the precursors of melanocytes, migrate along the dorsolateral pathway in chick and mouse. In zebrafish, however, the migration pattern is more complex, with melanoblasts migrating along the dorsolateral and medial pathways, whereas iridophores and xanthophores use the medial and dorsolateral pathways respectively (Figure 4).

Certain genes playing a key role in specification of melanocytes from the neural crest have been well-characterised. At the core of melanocyte fate-specification is the transcriptional activator microphthalmia-associated transcription factor (MITF; Mitfa in zebrafish). This conserved transcription factor is a member of the basic helix-loop-helix leucine zipper (bHLH-LZ) family, members of which possess DNA-binding and dimerization domains. MITF has an amino-terminal transactivation domain which triggers transcription and a domain mediating dimerisation (HLH-LZ) in the carboxy-terminal region. It is expressed in all melanocytes throughout their development and maintained in the differentiated cell (Lister *et al.*, 1999; Steingrímsson *et al.*, 2004; Hodgkinson *et al.*, 1993; Tachibana *et al.*, 1994). In mouse Mitf and zebrafish *mitfa* null mutants, all melanocyte markers are absent in embryos and adults, except for transient expression of Mitf/*mitfa* itself (Opdecamp *et al.*, 1997; Lister *et al.*, 1999; Greenhill *et al.*, 2011). In humans, MITF mutations cause two distinct syndromes:

Tietz syndrome (OMIM #103500) and Waardenburg syndrome type 2A (WS2A; OMIM #193510). Tietz syndrome, also called Tietz albinism-deafness syndrome, is characterised by loss of pigmentation (without affecting the eyes) and complete sensorineural hearing loss (Levy *et al.*, 2006). WS2A is characterised by complete absence of pigmentation in patches of skin and hair, especially in the ventral midline and forelock and sensorineural hearing loss (Opdecamp *et al.*, 1997). Mitf mutations also cause forms of albinism (including ocular albinism) and sensorineural deafness (OMIM #103470). Finally, Mitf mutation (amino acid substitution E318K) increases the susceptibility to cutaneous malignant melanoma (OMIM #614456). Mitf expression is driven by several transcription factors including Sox10, Pax3, CREB and the Wnt pathway. The protein is required for initial melanoblast specification, melanocyte differentiation, and has an ongoing role in melanocyte maintenance (Lister *et al.*, 1999; Johnson *et al.*, 2011). Sox10 directly regulates the activity of Mitf. In zebrafish *mitfa* gene promoter contained several Sox10 protein binding sites which bind directly Sox10 and were necessary for adequate Mitf expression (Elworthy *et al.*, 2003). PAX3 has been shown to regulate MITF expression by binding to the MITF gene promoter in studies performed using 624-mel cells cultures (Watanabe *et al.*, 1998), and HeLa cells (Bondurand *et al.*, 2000). Lang *et al.* proposed that Pax3 cooperated with Sox10 to activate Mitf expression, while it simultaneously prevented Mitf from activating downstream genes (Lang *et al.*, 2005). Activated CREB binds the cAMP response element consensus motif, which is located in the MITF promoter to increase the MITF gene expression (Bertolotto *et al.*, 1998). Moreover, both α-MSH (Melanocyte-stimulating hormone) and ultraviolet can initiate the activation of CREB, which is followed by the up-regulation of MITF (Luger *et al.*, 1997). LEF-1 is a transcription factor that is involved in the Wnt signal transduction pathway. Saito *et al.* proposed that MITF expression in melanocytes was triggered by the Wnt signalling pathway through LEF-1 and temporally benefited from the functional cooperation of LEF-1 and MITF (Saito *et al.*, 2002).

Work in mouse and zebrafish indicates that SOX10 drives Mitf/*mitfa* transcription from the very earliest stages, making it an early marker of neural crest cells (Dutton *et al.*, 2001; Elworthy *et al.*, 2003). Curran *et al.* showed that most of the zebrafish

NCC transiently turn on *mitfa*. However, only around half of them activate expression of melanocyte markers such as *dct*; other cells switch to other fates, including iridophores (Curran *et al.*, 2010). Forced expression of *mitfa* in zebrafish *sox10* or *mitfa* mutants, rescues the melanocyte phenotype equally well, strongly suggesting that the primary role of Sox10 in melanocyte development is to activate *mitfa* expression in melanocyte progenitors (i.e., to specify melanoblasts) (Elworthy *et al.*, 2003). However, in mammals, PAX3 has also been shown to have a role in the early specification of melanoblasts. SOX10 and PAX3 have been suggested to promote synergistically the activation of the Mitf promoter. Moreover, PAX3 can also directly activate the Mitf promoter. Hence, PAX3 is proposed to enhance the specificity and effects of SOX10 (Watanabe *et al.*, 1998; Bondurand *et al.*, 2000; Potterf *et al.*, 2000).

Wnt signalling also plays a role in melanocyte specification. An elevated Wnt signalling biases cranial NCC to adopt a melanocyte fate (Dorsky *et al.*, 1998). Wnt-induced signaling is mediated by Tcf/Lef transcription factors, which binds directly to the *mitfa* promoter (Dorsky *et al.*, 2000).

Choosing between alternative fates is a fundamental problem in stem cell biology. In mouse, melanocytes may be derived from bipotent Schwann cell progenitors (SCPs), and hence these cells choose between melanocyte and the alternative Schwann cell fate. Although much remains unknown regarding the molecular basis of this fate decision, current data identifies four factor interactions that may be important: First, melanocyte formation from SCPs is inhibited by ErbB signaling. Secondly, in mouse and chick, SOX2 represses Mitf expression and thus inhibits melanocyte formation. Conversely, MITF represses Sox2 expression, hence these mutually repressive interactions are important for SCPs' commitment to either Schwann cell fate or melanocyte lineages (Adameyko *et al.*, 2009). Finally, in avian and murine SCPs, loss of Foxd3 function induces differentiation into melanocytes, suggesting that FOXD3 suppresses a melanocyte fate, while promoting Schwann cell and glial fates (Nitzan *et al.*, 2013).

There are several features of the melanocyte GRN in zebrafish that play a role in melanocyte commitment (Greenhill *et al.*, 2011). Sox10 maintains multipotency (Kelsh, 2006), and its loss would be expected to contribute to commitment. Mitfa

and Sox10 initially establish a positive feedback loop that promotes melanoblast specification, but subsequently, Mitfa-dependent activation of the Hdac1 complex, represses *sox10* in melanoblasts promoting both differentiation and fate commitment (Greenhill *et al.* 2011).

Two key signalling factors play crucial roles in melanocyte migration. First, endothelins (ET) are ligands for seven pass G-protein coupled endothelin receptors (Ednr) and have been found to be expressed in early melanoblasts. A number of studies suggest that ET3/EDNRB2 signalling is crucial for melanoblast migration (Lecoin *et al.*, 1998). Mutations in the human EDNRB gene lead to Waardenburg syndrome type 4A, an auditory-pigmentary syndrome characterised by patchy hypopigmentation, congenital sensorineural hearing loss, and Hirschsprung disease (OMIM #277580). Interestingly, Et3 has also been reported to drive proliferation of melanocytes at least in cell culture, suggesting that Ednrb signalling may have multiple functions (Lahav *et al.*, 1996; Reid *et al.*, 1996). In zebrafish, ednrb1a has a clear role in adult pigment pattern formation. However, although ednrb1a is expressed in early pigment cells, there are no pigment pattern defects in ednrb1a mutant embryos (Parichy, Ransom, *et al.*, 2000).

In mice, the major signalling molecule promoting melanocyte migration has been shown to be KIT. KIT is a type III receptor tyrosine kinase that responds to Stem Cell Factor (SCF, also called Steel factor) binding by inducing PI(3) kinase and MAP kinase cascades. KIT has multiple conserved roles in melanocyte development, including melanoblast survival, proliferation and appropriate dispersal along the dorsolateral migratory pathway. First, Kit is expressed from premigratory stages in melanoblast development in mouse and zebrafish (Wehrle-Haller *et al* 1995, Parichy *et al* 1999). Second, Kit and SCF mutant mice and zebrafish display reduced numbers of melanocytes, with severe deficits in ventral regions (Alexeev and Yoon, 2006), suggesting reduced migratory capacity of melanocyte precursors. Kit signalling was proposed to initiate melanoblast migration, attracting them onto the dorsolateral pathway lined by cells expressing SCF. In zebrafish *kit* mutants, melanocytes are specified and differentiate, but they fail to migrate properly and tend to accumulate in premigratory positions (dorsal to the neural tube, and in a cluster posterior to the

inner ear) (Parichy *et al.*, 1999). This second role of KIT is proposed to be a chemokinetic effect: KIT increases the melanoblasts' rate of movement and, therefore, their chance of encountering the follicle (Jordan and Jackson, 2000).

There needs to be an increase in cell numbers to produce the full complement of skin pigment cells from the limited number of NCC, so it is clear that proliferation and survival of melanocyte precursors are important. In zebrafish and mammals, survival and proliferation of NC-derived melanoblasts appears to be critically dependent on Kit signalling. In fish, a temperature-sensitive conditional *kit* allele has enabled demonstration that Kit's role in survival begins from around 2 dpf and continues until approximately 4 dpf (Rawls and Johnson, 2003). Melanoblast apoptosis in *kit* mutants results from a failure of the receptor tyrosine kinase to activate the PI(3) kinase and MAP kinase cascades. PI(3) kinase delivers a powerful anti-apoptotic signal via the activation of the kinases PDK1 and AKT (Goding, 2000). Another target of Kit signaling is Bcl2, a known pro-survival gene. Disruption of Bcl2 function in melanocytes results in their apoptotic loss in mice and in cell culture (McGill *et al.*, 2002). Cell-cycle exit, and thus the termination of proliferation, appears to also be MITF-dependent, and results from activation of *p21* (WAF1) expression, a cyclin-dependent kinase inhibitor (Carreira *et al.*, 2005).

The melanosome is a specialised membrane-bound organelle, which is involved in the synthesis, storage and transport of melanin. Melanosome organellogenesis progresses through four stages of maturation involving multiple enzymatic and structural proteins. Many proteins have been identified as melanosome-specific, including tyrosinase (TYR), tyrosinase-related protein 1 (TRP1), dopachrome tautomerase (DCT), ocular albinism type 1 protein (OA1), melanoma-associated antigen recognised by T cells (MART-1), vesicle amine transport protein 1 homolog (VAT-1), oculospanin, syntenin, and glycoprotein non-metastatic melanoma protein b (GPNMB) (Hoashi *et al.*, 2010). Their roles are diverse, including melanin synthesis, regulation of ion transport across the melanosome membrane and structural functions, which we will briefly touch upon here.

Melanocytes are highly dendritic cells that reside in the basal layer of the epidermis once they have migrated. In mammals, dendrites are used to transport melanosomes

from the cell body to the neighbouring keratinocytes. In contrast, in fish, amphibians and reptiles, melanosomes and pteridine granules are not secreted, but they are often motile, allowing for changes in coloration and the skin pigment pattern. Dendrite formation, maintenance and activity ae reliant upon a specialised cytoskeleton (Lacour *et al.*, 1992). Ednrb signalling promotes dendricity in zebrafish melanocytes (Parichy *et al.* 2000).

Actin filaments are the main structural component of dendrites and they play a central role in melanosome transport. This role may be mediated by actin binding proteins like myosin Va, melanophilin, RAB27A, and SLP2-A, all shown to transport melanosomes (Matesic *et al.*, 2001; X. Wu *et al.*, 2002; X. S. Wu *et al.*, 2002; Fukuda *et al.*, 2002). The currently accepted model describes that RAB27A recruits melanophilin to the melanosome membrane, which allows the recruitment of myosin Va to link the organelle to the actin cytoskeleton (Hume and Seabra, 2011).

Melanin is produced in the melanosomes by the cooperative action of enzymes of the tyrosinase family. Tyrosinase is considered as the key enzyme in this process. It is a membrane glycoprotein with a single transmembrane helix. The C-terminal sits in the melanocyte cytosol and the N terminal is located inside the melanosome lumen. The enzyme is activated by cytosolic protein kinase C (PKCβ) (Hearing and Jiménez, 1987). Tyrosinase can use both L-tyrosine and L-DOPA, which is formed in melanosomes from L-tyrosine by the enzyme tyrosine hydroxylase isoform I (THI), as substrates (Marles *et al.*, 2003). The concentration of L-tyrosine in the melanosome depends on the conversion of the amino acid L-phenylalanine by phenylalanine hydroxylase (PAH) (Schallreuter *et al.*, 1994). Therefore, all the three enzymes PAH, THI and tyrosinase are crucial for the initial steps of the melanogenesis. The activities of PAH, THI and tyrosinase are controlled by the cofactor 6BH4 which acts as the essential electron donor for PAH to produce L-tyrosine from L-phenylalanine and for THI to convert L-tyrosine to L-DOPA (Schallreuter *et al.*, 2008). Tyrosinase converts tyrosine to DOPAquinine and DOPAchrome that will be used by other melanogenic enzymes such as DCT or TYRP1 to produce eumelanin or pheomelanin (Hearing, 2011). Mitf, the melanocyte main transcription factor, drives the expression of these

melanogenic enzymes through the M-box or E-box that they contain in their promoter (Hou *et al.*, 2006).

The pigment-type switching system that controls whether melanocytes produce black/brown eumelanin or yellow/red pheomelanin, is regulated by the melanocortin 1 receptor (MC1R), a G-protein-coupled receptor expressed in melanocytes. It involves the action of α-melanocyte-stimulating hormone (α-MSH) and the opposing agouti signal protein (ASP) (Walker and Gunn, 2010). MC1R activity is sufficient to produce eumelanin in its basal state, but production increases upon α-MSH stimulation. The increased MC1R signalling results in the activation of Mitf, and then of the melanogenic enzymes (e.g., TYR, TYRP1, DCT), resulting in the production of more eumelanin (García-Borrón *et al.*, 2005). ASP competes with α-MSH for binding to MC1R. When MC1R is inactivated by ASP pheomelanin is produced instead (Rouzaud *et al.*, 2003). Fish melanophores contain only eumelanin (as they cannot produce pheomelanin). Melanocortins are involved in the melanocyte production in fish, probably through its binding to Mc1R leading to the upregulation of the melanogenic genes (Logan *et al.*, 2006). Recent studies suggest that agouti may also have a role in fish melanogenesis (Guillot *et al.*, 2012).

All the information above suggests that melanocyte development is a very complex process, regulated by many different genes, with several different functions (some are transcription factors, some are membrane receptors, and others are enzymes). All these genes interact in a dynamic way to form an intricate gene regulatory network that drives melanocyte development.

Introduction to gene regulatory networks

To understand the behaviour of the complex interactions that occur in gene regulatory networks and their function a mathematical modelling approach might be useful and informative. Intuition alone is not enough to comprehend the effects of multiple simultaneous interactions within these networks and their dynamical evolution. Mathematical models are particularly useful to understand the effects of

nonlinear interaction such as those that describe the oligomerisation of subunits to form a functional TF, as well as to predict the genetic response of normal and mutant organisms (Smolen *et al.*, 2000).

There are two main approaches to build a GRN model: the logical-network or "Boolean" method and the dynamic-systems one. The former takes into account only the simple details: the expression of each gene is *modelled as* either on or off, there is no intermediate level. These simple models allow use of large time steps and minimise computational time required for simulations. The dynamic-systems approach uses ordinary differential equations to describe the rate of change of the concentration of mRNA or proteins. The terms of the equations describe how the gene expression is modified by changes in the effector molecule(s) (transcription factor(s)). Where appropriate, stochastic fluctuations in molecule numbers due to random events in its synthesis or degradation can be incorporated to the model. This method is usually preferred over the logical-network one because it is more accurate, since gene expression levels are continuous variables rather than ideal "on or off" switches. However, dynamic-systems methods use much shorter time steps and as a consequence they require much more computer time. Also we do not always know enough about gene concentrations and how they change. That is why sometimes logical networks represent the only practical alternative to model lengthy processes such as development of tissues (Smolen *et al.*, 2000).

Gene regulatory networks have been successfully used to describe biological processes in several organisms and tissues: in the yeast *Saccharomyces cerevisiae* (Alcasabas *et al.*, 2013), in the mammalian immune system (Singh *et al.*, 2014), in the plant *Arabidopsis thaliana* (Wellmer *et al.*, 2006), in the sea urchin (Davidson *et al.*, 2002) or the fly *Drosophila* (Aguilar-Hidalgo *et al.*, 2013) just to name a few. However, the combined use of both GRN and mathematical modelling to investigate biological processes is not as extended. One of the most studied GRNs is the one that describes the sea urchin development. The Davidson lab have been working on this GRN for many years, and they have developed a very complete and detailed network of this organism. Part of their work was focused on the development of a gene network that explains how endomesoderm specification in sea urchin is controlled (Davidson *et*

al., 2002). Another group used this GRN to create a mathematical model that describes endomesoderm specification, which allowed them to predict new interactions and genes that where initially located in erroneous positions; therefore it contributed to the improvement of the model (Kühn *et al.*, 2009). Mathematical modelling was also used to understand the GRN underlying the development of the *Drosophila* ocellar complex. The model allowed them to expand the GRN to incorporate new genes that explain how the interaction between Hedgehog signalling pathway and the GRN results in stable gene expression that allows ocellar development (Aguilar-Hidalgo *et al.*, 2013). Hence, mathematical modelling of GRN have proven to be very useful to understand better how a GRN works and to improve them.

Gene regulatory network of melanocyte development

Greenhill *et al.* (2011) used a systems biology approach to identify and develop the GRN underlying melanocyte specification and differentiation in zebrafish. They used a mathematical modelling approach in conjunction with experimental data in an iterative process to create a model that makes specific predictions about the properties of unidentified factors in melanocyte differentiation and explains the melanocyte development in a semi-quantitative manner.

They constructed a dynamical model of the GRN based on ordinary differential equations, considering the transcription levels as dynamic variables. The external activatory signal in the network was designed as Factor A, comprising several transcription factors whose function in the regulation of the network is poorly understood: Lef/Tcf (Wnt pathway), Sox9, FoxD3, Pax and AP2. These input factors regulate Sox10 which had been shown to activate the expression of Mitfa directly (Elworthy *et al* 2003), termed the master regulator of melanocyte specification, which will activate the melanogenic genes (Dct, Tyr, Tyrp1 and Si) as well as the *sox10* promoter. Subsequently, Sox10 represses the expression of the melanogenic genes and its expression is repressed by the recruitment to the *sox10* promoter an Hdac1

complex that deacetylates its chromatin and represses *sox10* transcription by Mitfa. In that way, Factor A-dependent *sox10* expression as well as Mitfa-dependent expression will both be inactivated. The model proposed also the existence of a Sox10-independent, Mitf-dependent transcriptional activator of *mitfa* (Factor Y) that provides a positive feedback loop to allow stable melanocyte differentiation. This feature was not initially part of the model, but was incorporated because the model simulations highlighted the need of this factor to guarantee stable melanocyte development. This was also supported by experimental data that showed a reduction of *mitfa* expression in *mitfa* mutant, due to the absence of the positive feedback through Factor Y. Finally, Sox9b drives Sox10 and Mitfa-independent melanisation and it is activated by Factor B, which has a transient role and is restricted to the early phase of melanocyte development (Greenhill *et al.*, 2011) (Figure 5).

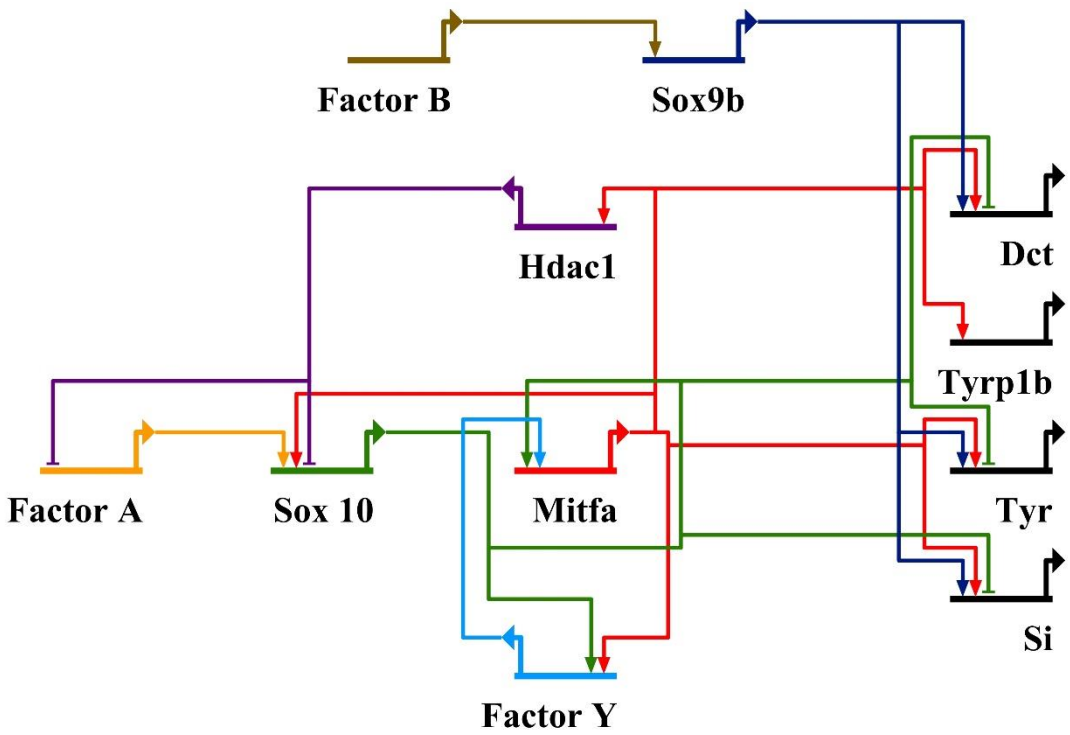

Figure 5. Core GRN driving melanocyte differentiation proposed by Greenhill *et al.* (2011)

The mathematical model was constructed as a one stage process: binding and unbinding of TF to DNA regulate protein production in a single step of synthesis,

without taking into account mRNA levels. The model was produced using ordinary differential equations. TF binding and unbinding to the DNA was described as a faster process than protein synthesis and degradation, in that way the dynamics were solved in conditions of quasi-equilibrium for the TFs. The result was a description of both activatory and repressive regulations in terms of Hill-like functions (Greenhill *et al.* 2011). The equations that make up the model have the following structure:

$$\frac{dA}{dt} = g\alpha B - dA$$

Where A is the TF concentration, *g* is the rate of TF synthesis, B is the concentration of the regulatory TF, *d* is the rate of TF degradation and α is a factor that measures the rate for binding and unbinding of the regulatory TF i.e. B. The whole set of ordinary differential equations that form the model can be found in Appendix 1.

Greenhill *et al.* (2011) developed the current GRN model by expanding and refining an initial melanocyte GRN and model in an iterative approach to obtain the model that best described the data they had. Thanks to the use of this systems biology approach they proposed new network players and interactions which were not present in the initial model, including the Sox10-mediated repression of many Mitfa target genes; the transient expression of Sox10 in differentiating melanocytes as a consequence of the Mitfa-dependent repression of *sox10* through a mechanism involving Hdac1 complex; a Sox10 independent weak activation of melanogenesis genes; and that *mitfa* expression is directly or indirectly, Mitfa-dependent.

This model allowed them to predict the evolution of gene expression of the GRN genes along the first 100 h of zebrafish embryonic development. Figure 6 represents the changes on expression levels of the different genes of the network along time. However, the data they have was very limited: *in situ* hybridisation expression patterns that can be used to get coarse expression data. Therefore, when they had to find the parameters of the model they chose the ones that produced gene expression curves that were similar in shape to what they expected to happen based on the *in situ* data. That means that the quantities and exact detail of the curves are not exact and may be somehow erroneous.

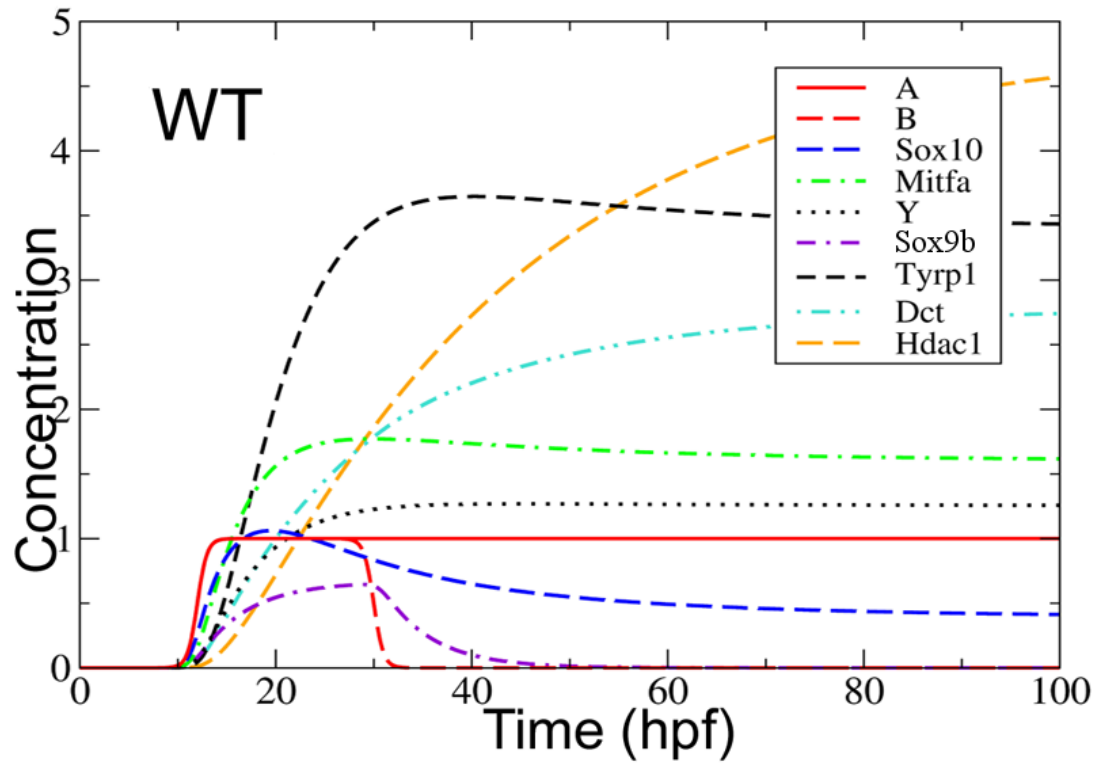

Figure 6. Evolution of gene expression in the GRN over time. The graphic shows that the input factors (A and B) start acting at 12 hpf. Factor B is only transient and it ends at 24 hpf, while factor A keeps acting during all the simulation. The rest of the curves represent the evolution of gene expression levels of the genes in the network (Greenhill *et al.* 2011).

This model represented a big improvement on the information previously available on the melanocyte GRN, and helped understand much better how these genes interact to drive melanocyte development. However, in order to obtain a model that is closer to reality we need to base it on quantitative data. Thus in this study we aimed to obtain quantitative data from some genes of the GRN using qPCR to develop a quantitative mathematical model, that represents better the reality. At the same time this network includes some of the main genes involved in melanocyte development, but there are more important genes involved that are not currently described. Hence in this work we intend to identify novel genes involved in this process that should be part of the melanocyte GRN.

Aim and objectives

The aim of this study is to expand the GRN including more genes involved in melanocyte development and to develop a quantitative mathematical model for the GRN by:

- Identifying novel genes involved in the melanocyte gene regulatory network and their function.
- Understanding how these novel genes interact with the current genes that form the core melanocyte GRN and finding their place in the network.
- Obtaining quantitative data from some of the main gene in the melanocyte GRN in several time points throughout melanocyte development.
- Developing a quantitative mathematical model of the GRN using the obtained data to carry out a parameter optimisation procedure.

Chapter 2:
Materials and Methods

Materials

Chemical Reagents

UltraPure Agarose	Invitrogen Life Technologies, UK
Low Melting Agarose	Invitrogen Life Technologies, UK
Phenol-Chloroform-isoamyl alcohol	Sigma Aldrich, Germany
Tricaine (Ethyl 3-aminobenzoate methanesulphonate)	Sigma Aldrich, Germany
Tween 20	Sigma Aldrich, Germany
Luria agar	Sigma Aldrich, Germany
Luria broth	Sigma Aldrich, Germany
Formaldehyde	Sigma Aldrich, Germany
Ethanol	Sigma Aldrich, Germany
Methanol	Sigma Aldrich, Germany
PBS	Fisher Scientific, UK

Kits

DIG RNA labelling *kit* SP6/T7	Roche Diagnostics, Germany
NycleoBond Xtra Midi	Macherey-Nagel, Germany
Zymoclean Gel DNA Recovery *kit*	Zymo Research, USA
DNA Clean & Concentrator	Zymo Research, USA
GeneJET Plasmid Miniprep Kit	Life Technologies, UK
Zero Blunt TOPO PCR Cloning Kit	Life Technologies, UK
Experion RNA StdSens Analysis Kit	Bio-Rad, USA

geNorm 12 gene *kit* for zebrafish	PrimerDesign, UK
TSA Plus Cyanine 3/Fluorescein System	PerkinElmer, USA

Antibiotics, Indicators and Dyes

Ampicillin	Sigma Aldrich, Germany
Ethidium Bromide	Sigma Aldrich, Germany
Kanamycin	Sigma Aldrich, Germany
Fast Red tablets	Roche Diagnostics, Germany
NBC/BCIP	Roche Diagnostics, Germany

Enzymes

Proteinase K	Fisher Scientific, UK
RNase A	Promega, UK
RNase inhibitor	Promega, UK
Pronase	Sigma Aldrich, Germany
T4 ligase	Promega, UK
Go Taq polymerase	Promega, UK
KOD Hot Start DNA polymerase	Novagen, USA
SuperScript III Reverse Transcriptase	Invitrogen Life Technologies, UK
RNase-Free DNase	Promega, UK
Precision DNase	PrimerDesign, UK
nanoScript 2 Reverse Transcriptase	PrimerDesign, UK
PrecisionPLUS Mastermix	PrimerDesign, UK

Nucleic acids

dNTP mix	Promega, UK

DIG labelled dNTPs	Roche Diagnostics, Germany
DNA ladder 100 bp	New England BioLabs, UK
DNA ladder 1kb plus	Thermo Scientific, USA

Solutions and Buffers

Alkaline phosphate buffer

100 mM Tris-HCl pH 9.5

50 mM MgCl2

100 mM NaCl

0.1% (v/v) Tween20

Blocking reagent stock solution

10 % (w/v) of Blocking Reagent (Roche) in 1xMAB can be prepared and kept in the freezer.

Embryo medium (50x stock solution)

0.5 µM NaCl

0.17 µM KCl

0.33 µM CaCl

0.33 µM MgSO4

0.1 % methylene blue

In distilled water

Hybridisation mixture (HM)

50 % (v/v) formamide

5x SSC

50 ng/mL heparin

500 ng/mL tRNA

0.1% (v/v) Tween20

9.2 mM citric acid

In DEPC-treated Milli-Q water

Luria Agar

3.7 % (w/v) Luria agar base in distilled water. Autoclave.

Luria broth

2.5 % (w/v) Luria broth base in distilled water. Autoclave.

MAB 5x

500 mM maleic acid

450 mM NaCl

In Milli-Q water, adjust pH to 7.5 and autoclave

MABT

Dilute MAB 5x with Milli-Q water

Add 0.1% (v/v) Tween20

Phosphate buffered saline Tween (PBST)

0.1% (v/v) Tween20 in PBS

Paraformaldehyde (PFA)

4% (w/v) paraformaldehyde in PBS

Autoclaved PBS is firstly warmed to 60^{o}C and then PFA is added, pH is adjusted to 7.0 to improve PFA solubilisation, solution should never be warmed above 60^{o}C and should be frozen in aliquots as soon as cleared to prevent degradation.

Pronase

20 mg/mL stock solution of Pronase (Protease from *Streptomyces griseus*, Sigma-Aldrich) in PBS, dilute to 2 mg/mL in EM prior to use.

SSC 20x

3 M NaCl

300 mM sodium citrate

In Milli-Q water adjusting pH to 7.0 with 1 M HCl

SOB (Super optimal broth)

0.5% (w/v) yeast extract

2% (w/v) tryptone

10 mM NaCl

2.5 mM KCl

20 mM MgSO4

Per litre:

5 g yeast extract

20 g tryptone

0.584 g NaCl

0.186 g KCl

2.4 g MgSO4

Adjust to pH 7.5 prior to use.

SOC (Super optimal broth with catabolite repression)

1 L SOB

20 mM glucose solution (20 ml)

TE 1x

10 mM Tris-HCl (pH 8.0)

1 mM EDTA (pH 8.0)

10 mM NaCl

TAE 1x

40 mM Tris-HCl

20 mM sodium acetate

2 mM EDTA

In Milli-Q water adjusting the pH to 7.8 with glacial acetic acid

Tricaine stock solution (3-amino benzoic acidethylester)

400 mg tricaine powder

87.9 ml dd water

2.1 ml 1M Tris (pH 9). Adjust pH to 7.

To use tricaine as anaesthetic, dilute 4.2 ml of tricaine solution in 100 ml of water

Primers

Primers were purchased either from SourceBioscience, Invitrogen, Eurofins or PrimerDesign and are listed in table X.

Table 1. Primers. The sequences are given in the 5' to 3' direction of their respective DNA strands.

Primer name	**Primer sequence**
Sox10 Sense (PrimerDesign)	GCATCCATCCTTCCTGAACT
Sox10 Anti-sense (PrimerDesign)	AAAGATTATTCATTCATACAAATACAGA
Kita Sense (PrimerDesign)	GCGATTTGGAAAAACTCTTGGAT
Kita Anti-sense (PrimerDesign)	GCACTCGGTTTCAGCATCTT
Dct Sense (PrimerDesign)	CAGTGACCCATCCCGAATC
Dct Anti-sense (PrimerDesign)	CCTCATTGGTGCCGTTACA
Mitfa Forward	CTGGACCATGTGGCAAGTTT
Mitfa Reverse	TGAGGTTGTGGTTGTCCTTCT

Methods

Zebrafish techniques

FISH HUSBANDRY

Zebrafish and zebrafish embryos were manipulated in accordance with the Home Office regulations and as suggested in "The zebrafish book" (Westerfield, 2000). Fish stocks were maintained in the University of Bath zebrafish facility.

Embryos were obtained from crosses of wild type AB fish and several mutants: *sox10*^{m618}, *sox10*^{t3} (Dutton *et al.*, 2001), *mitfa*^{w2} (Lister *et al.*, 1999)and *kit*^{j1e249} (Rawls

and Johnson, 2003). Embryos were raised and staged according to Kimmel *et al.* (1995). Embryos were grown in embryo medium using 90mm Petri dishes in a MIR-155 incubator (Sanyo) at 28.5 ^{o}C, unless stated otherwise.

PTU TREATMENT

To prevent melanin synthesis to achieve better internal imaging of the embryos, they were incubated in EM with 0.003% of PTU (1-phenyl-2-thiourea) (Sigma-Aldrich) from 24 hpf onwards until embryos were fixed.

DECHORIONATION

Embryos were dechorionated using either the manual method or the enzymatic one. Manual dechorionation involved using two Watchmakers' forceps No. 5. One of them was used to hold the embryo and the other one to pull apart the chorion.

The enzymatic dechorionation involved the use of Pronase. Embryos were incubated with 1x Pronase in EM at 28.5^{o}C during some minutes, until the embryos are outside the chorion. Pronase was then removed and embryos were rinsed several times with fresh EM.

ANESTHETISING EMBRYOS

Embryos older than 15 hpf were anesthetised with tricane (ethyl 3-aminobenzoate methanesulfate salt) to approximately 0.2% of the final volume prior to manipulation.

DISSECTION OF EMBRYO TRUNKS FOR QUANTIFICATION EXPERIMENT

A pool of 40 embryos were transferred to a plate with fresh PBS where there were manually dechorionated and then anesthetised with tricane. 5 of these embryos were transferred with a small amount of PBS to a coverslip and their trunks were cut with dissecting pins under a dissecting scope. The fragment of the fish dissected goes from the beginning to the end of the yolk extension in all the stages. The 5 trunks were collected with a micropipette taking as little PBS as possible and deposited in a PCR tube. The tube was spun down and as much liquid as possible was removed without taking out the trunks. 50 µl of QIAzol was added, tube was vortexed for a

couple of minutes and then stored at -80°C until the RNA extraction was performed. The remaining 35 embryos were collected into tubes and fixed with 4% PFA and stored at 4°C to be used in ISH.

Molecular Methods

BACTERIAL GROWTH

Glycerol stocks or agar stabs of bacterial stocks were propagated by streaking on LB-agar plates containing appropriate selective antibiotics and incubated overnight at 37°C.

Bacterial cultures were grown in LB-medium supplemented with the required antibiotics at the following concentrations: 50 µg/ml of ampicillin or 25 µg/ml of kanamycin. After inoculation with a single colony, the culture was incubated overnight at 37°C in a shaking incubator.

TRANSFORMATION OF BACTERIAL CELLS

The transformations of plasmid DNA into E. coli were performed by heat shock. During the entire project DH5αF' (Clontech Laboratories) strain was used; these were prepared in the lab using chemically induced competency and aliquoted in 1.5 mL centrifuge tubes (following the protocol from Dr. K Kaji lab). Aliquots were defrosted on ice for 20 min with occasional tapping of the tube. 5 µl of ligation or 1 µl of a stock plasmid were added and the cells were left on ice for 30 min with occasional gentle mixing. Tubes were incubated at 42°C for 45 s and immediately placed on ice and left for 2 min. After transformation, cells were incubated in 1 ml of SOC medium for one hour at 37°C with gentle shaking before streaking 200 µl on LB-agar plates containing the appropriate antibiotic. Plates were incubated overnight at 37°C and positive colonies were then picked and grown in LB-medium with the appropriate antibiotic at 37°C overnight in a shaking incubator. The culture volume was chosen according to the specific extraction protocol.

PHENOL/CHLOROFORM/ISOAMYL EXTRACTION AND ALCOHOL PRECIPITATION OF DNA

For the extraction of RNA or DNA, Milli-Q water was added to the sample of dirty DNA to 100 µl, followed by 100 µl of phenol/chloroform/isoamyl alcohol pH 8 (Sigma-Aldrich). Sample was vortexed for 15 s and centrifuged for 3 min at maximum speed. Supernatant was recovered and 2.5 volumes of ethanol and 1/10 volume of NaOAc (3 M, pH 5.4) were added. Sample was then precipitated overnight at -20^{o}C or 30 min at -80^{o}C. Sample was centrifuged at maximum speed for 30 min at 4^{o}C, supernatant was decanted and pellet rinsed with 70% ethanol. Pellet was dried for 10 min at room temperature and re-suspended in water.

PLASMID DIGESTION

Plasmid digestion by restriction enzymes was performed in a minimal volume, using the conditions recommended by the manufacturer. When required BSA was added and heat inactivation at 65^{o}C was carried out unless the reaction was used to run on an agarose gel or for Phenol/Chloroform precipitation.

IN SITU HYBRIDISATION PROBE SYNTHESIS

Prior to probe synthesis 10 µg of midi-prep plasmid DNA containing the cDNA of interest (template) was linearised by digesting the 3' end of the antisense sequence. The reaction was cleaned by phenol/chloroform/isoamyl alcohol extraction followed by alcohol precipitation.

For *in situ* hybridisation a RNA probe labelled with digoxigenin was created using the DIG RNA Labelling Kit (Roche, Cat# 11175025810). Probe was synthesised in an RNAse-free tube with the following components: 1 µg of the purified template, 2 µl of dNTP labelling mix, 2 µl of 10x transcription buffer, 1 µl of RNase inhibitor, 2 µl of 100 mM DTT, 2 µl of the appropriate RNA polymerase (SP6, T7 or T3) and DEPC water to a final volume of 20 µl. This was mixed gently and spun to the bottom of the tube before placing the tube in a 37^{o}C incubator for 2 – 4 hours. The cDNA template was digested by adding 2 µl of RNase-free DNase incubating it at 37^{o}C for 15 min. The reaction was stopped and the probe was precipitated by adding 2 µl of 0.5 M EDTA

pH 8, 2.5 µl of 4 M LiCl and 75 µl of 100% ethanol. The mixture was placed in -80°C for 30 min and then it was centrifuged at maximum speed for 30 min at 4°C. Supernatant was removed and the pellet washed with 70% ethanol centrifuging it for 5 min. The supernatant was discarded and pellet was dried briefly and resuspended in 100 µl of Milli-Q water.

PROBE FRACTIONATION

Alkaline hydrolysis was used to fragment a probe in order to increase its permeability inside the embryo, following the Parichy laboratory protocol. This is generally recommended for probes longer than 300 nucleotides. The following components had to be mixed: 50 µl of the intact probe, 30 µl of DEPC water, 10 µl of 0.4 M NaHCO3, 10 µl of Na2CO3. This reaction had to be heated in a water bath at 60°C for the time calculated by the following equation:

Time (min) = (starting kb – desired kb) / (0.11 x starting kb x desired kb)

Hydrolysed probe was then precipitated by mixing 100 µl of the fractionated probe, 100 µl of DEPC water, 20 µl of 3 M NaOAc pH 4.5, 2.6 µl of glacial acetic acid and 600 µl of ethanol. Mixture was precipitated for 30 min at -80°C and then centrifuged at maximum speed for 30 min at 4°C. Supernatant was removed and dried briefly and the pellet was resuspended by vortexing in 50 µl of DEPC water.

RNA EXTRACTION FOR QPCR

The embryo trunks stored at -80°C in QIAzol were thawed in ice. Half the volume of chloroform was added (i.e. 25 µl) to the tube and then thoroughly vortexed for 15 s. The tubes were centrifuged for 10 min at 4°C at maximum speed. The aqueous phase (supernatant) was collected and transferred into a new tube. 2 µl of linear acrylamide was added, the tube was vortexed and an equal volume of isopropanol (around 40 µl) was added before mixing it with the vortex again. The samples were left resting at room temperature for 5 min before centrifuging them for 30 min at 4°C and max speed. The isopropanol was removed and 100 µl of 80% ethanol were added. The tubes were centrifuged again for 3 min at 4° and max speed and the supernatant was removed carefully to avoid losing the pellet. The pellet was dried briefly and then re-

dissolved in 12µl of water before measure the concentration using a Nanodrop spectrophotometer.

REVERSE-TRANSCRIPTION

The extracted RNA was reverse-transcribed to cDNA using the nanoScript2 kit from PrimerDesign following the manufacturer's instructions.

qPCR

The qPCR was run in 20 µl reaction containing 19 µl of the reaction mix and 1 µl of the cDNA template in 48-well plates. We prepared a reaction mix containing the PrecisionPLUS master mix (PrimerDesign), the appropriate primers (Table 1) and water. We put 19 µl of that mix into all the wells and 1 µl of the suitable template following the layout in Table 2.

Table 2. Layout of the qPCR plates followed in the quantification experiment. Sam means sample, Stcur means Standard Curve and NTC means no template control.

Sam1x1	Sam1x2	Sam1x3	Sam7x1	Sam7x2	Sam7x3	Stcur1x1	Stcur1x2
Sam2x1	Sam2x2	Sam2x3	Sam8x1	Sam8x2	Sam8x3	Stcur2x1	Stcur2x2
Sam3x1	Sam3x2	Sam3x3	Sam9x1	Sam9x2	Sam9x3	Stcur3x1	Stcur3x2
Sam4x1	Sam4x2	Sam4x3	Sam10x1	Sam10x2	Sam10x3	Stcur4x1	Stcur4x2
Sam5x1	Sam5x2	Sam5x3	Sam11x1	Sam11x2	Sam11x3	Stcur5x1	Stcur5x2
Sam6x1	Sam6x2	Sam6x3	Sam12x1	Sam12x2	Sam12x3	NTCx1	NTCx2

STANDARD CURVE

To produce the standard curve for the qPCR the plasmids containing the cDNA for the genes of interest were used (Table 3). In order to get these plasmids at high concentration a Midiprep was done (following the instructions of the kit manufacturer). Using the length of the insert and plasmid the copy numbers were calculated and the appropriate dilutions performed to obtain the 5 points needed for the qPCR experiment, ranging from 100,000 to 10 copies.

Table 3.Details of the plasmid containing our gene of interest amplicons that were used for the qPCR standard curves.

Gene name	Insert length	Plasmid backbone and length	Total length
dct	1800 bp	pBluescriptIISK, 3000 bp	4800 bp
mitfa	1600 bp	pBluescriptIISK, 3000 bp	4600 bp
sox10	2000 bp	pGEM-T Easy, 3015 bp	5015 bp
kit	3100 bp	pKRX, 2956 bp	6056 bp

AGAROSE GEL ELECTROPHORESIS

DNA products were normally visualised using agarose gel electrophoresis. Agarose gels 0.8-2% (w/v) in 1xTAE buffer were made by heating in a microwave with regular mixing until agarose was fully dissolved. Solution was allowed to cool down and ethidium bromide (10 mg/ml) was added to a final concentration of 0.4 mg/ml, mixture was then poured into an electrophoresis tray. Once solidified these were submerged with 1xTAE buffer in the electrophoresis tank. Samples were loaded into the gel with 1x loading dye and run at between 80-120V (PowerPac 300, Bio-Rad), depending on the purpose of the experiment and type of nucleic acid. The agarose gels were then visualised and photographed under UV light.

BLEACH GEL

This gel was used to check the integrity of RNA following the protocol described in Aranda *et al.* (2012). Agarose was measured accordingly to produce a 1% gel in 50 ml of 1xTAE buffer. 1 ml of commercial bleach was added and incubated for 5 min. The mixture was melted in the microwave and after adding the ethidium bromide the gel was casted and used in electrophoresis as usual.

Whole mount *in situ* hybridisation (ISH)

Dechorionated embryos were fixed in 4% PFA, 5 hours at room temperature or overnight at 4^{o}C. After fixation, embryos were rinsed in PBT and washed 3 times for 5 min in MeOH. Dehydrated embryos were stored at -20^{o}C in MeOH until needed but at least overnight.

The *in situ* protocol started with the rehydration of the embryos washing them for 5 min in a series of 75% MeOH/25% PBT, 50% MeOH/50% PBT and 25% MeOH/75% PBT, followed by five 5 min washes in PBT.

Embryos that already contain melanised cells were bleached to avoid the melanin interfering with the *in situ* staining. The bleaching process consisted in mixing 100 µl of 10% KOH, 50 µl of 30% H2O2 and 850 µl of Milli-Q water, transferring the embryos to a plate or well, and adding the mixture. Embryos were incubated in this mixture until they looked clear and the melanised cells were transparent; then embryos were transferred back to a tube and washed 3 times for 5 min in PBT.

Embryos were then incubated in Proteinase K (10 µg/mL) in PBT to permeabilise the cell walls and allow effective probe penetration. The incubation time depends on their developmental stage (Table 1).

Table 4. Proteinase K digestion times.

Developmental stage	Treatment time
Prior to 24 hpf	No treatment
24 hpf	5 min
26 hpf	5 min
30 hpf	8 min
32 hpf	8 min
36 hpf	10 min
48 hpf	12 min
60 hpf	12 min
72 hpf	15 min
96 hpf	20 min

After the treatment, embryos were fixed in 4% PFA for 1 hour at room temperature and then they were washed 5 times for 5 min in PBT. The previous steps (PFA fixation and PBT washes) were not necessary if the embryos were not digested with Proteinase K.

Embryos were then prehybridised for at least 1 hour at 68^{o}C in HM. Then the HM was removed and the embryos were incubated overnight at 68^{o}C with 200 µl of pre-warmed HM plus probe at 1/100 dilution.

The next day, embryos were washed for 5 min at 68°C in 66% HM/33% 2xSSC, 33% HM/66% 2xSSC and 2xSSC. Then, embryos were washed twice for 20 min in 0.1xSSC+0.1% Tween20 at 68°C. Afterwards, the samples were washed for 5 min at room temperature in 66% 0.2xSSC/33% PBT, 33% 0.2xSSC/66% PBT and PBT.

Blocking solution (diluting to 1% the stock solution) was added to the samples for 1 hour at room temperature. Then embryos were incubated for 2 hours at room temperature in anti-DIG-AP antibody (Roche) (1:5000 in 1% blocking solution). Finally embryos were washed 5 times for 15 min in PBT and left overnight at 4°C.

The following day, embryos were washed 4 times for 5 min in AP buffer before staining with NBT/BCIP solution or FastRed tablets following the manufacturer's guidelines (Roche). The staining reaction was carried out in the dark and was stopped when the signal was strong or background started to appear. The reaction was stopped by rinsing with PBT and fixing the embryos with 4% PFA for 2 hours at room temperature or overnight at 4°C. Finally, the embryos were transferred through a glycerol series of 30 %, 50% and 80% in PBS, before storing them at -20°.

Whole mount double fluorescence *in situ* hybridisation

The protocol was carried out as in non-fluorescence *in situ* hybridisation up to the step where the embryos were incubated with the antibody. In the case of a fluorescent *in situ*, instead of using anti-DIG-AP, the antibody used was anti-DIG-POD because the tyramide amplification reaction was going to be used in the next steps. The TSA System kit from PerkinElmer was used.

The anti-DIG-POD antibody was also pre-adsorbed and it was diluted 1:1000 in 1% blocking solution. After incubating the sample with the antibody for 2 hours at room temperature, the embryos were washed six times for 10 min each in PBT and they were stored in PBT overnight at 4°C.

The day after, the samples were incubated in amplification buffer for 10 min and then incubated for 1 hour protected from light in 50 µl of Cy3 tyramide working solution, made diluting Cy3 1:100 in 0.0015% H2O2/Amplification buffer. The embryos were

washed three times in PBT for 15 min and stored in SlowFade Antifade solution (Invitrogen) for detection under the microscope.

For a double fluorescence *in situ* hybridisation, two probes had to be hybridised: one labelled with DIG and the other with Flu (fluorescein). The protocol continued after the first fluorescent staining and instead of storing the sample in SlowFade Antifade solution, the anti-DIG-POD was stripped away incubating the sample in 0.1 M Glycine pH2.2 for 10 min.

The embryos were washed 5 times for 5 min in PBT and then incubated for one hour in 1% blocking solution. After that, pre-adsorbed anti-Flu-POD 1:500 in 1% blocking was added to the samples and they were incubated for 2 hours before washing 6 times for 10 min and storing the samples at 4^{o}C overnight.

Finally, the last day the samples were incubated in amplification buffer for 10 min before incubating them for 1 hour protected for light in 50 µl of FITC tyramide working solution, made diluting FITC 1:50 in 0.0015% H2O2/Amplification buffer. The samples were washed three times for 15 min in PBT and stored in SlowFade Antifade solution.

Embryo mounting and microscope techniques

Low power microscopic analysis and general embryo handling was performed under a Leica MZ12-FL dissecting microscope. Fixed embryos in glycerol were mounted in slide with a coverslip spacer and a coverslip on the top. Imaging was performed either on the Nikon Eclipse E800 microscope or the Zeiss Axio Imager.M2 microscope using the Apotome technology. Confocal microscopy was carried out using the Zeiss LSM510 META microscope.

Software

MICROARRAY ANALYSIS

The analysis of the microarrays Affymetrix Cell Intensity Files (.cel files) was performed using the Bioconductor package for the R statistical programming language. An R script was produced to perform this analysis (Appendix 3).

The ".cel" files were loaded into the workspace and then the data was pre-processed to eliminate the variation between arrays due to non-biological factors like sample processing. To get the differentially expressed genes the T-Test statistical test was used. The T-Test produced a list of genes ranked by t-statistics with the genes down-regulated in mutant embryos in the top of the list. The p-value is a measure of the probability of the test statistic being equal or greater than the observed result assuming that the null hypothesis is true. If the p-value is smaller or equal to the significance level then the null hypothesis (no difference between wild type and mutant embryos) can be rejected. In the T-Test all the genes with a p-value lower or equal than 0.05 were selected. When analysing microarray data thousands of test are conducted (one for each gene) leading to an accumulation of false positives. To address this, methods to correct multiple testing can be applied. In this analysis the False Discovery Rate (FDR) was applied. This method defines the proportion of false positives among all the genes identified as differentially expressed. At a defined FDR significance level (0.05) a known proportion of genes will be false positives and therefore these are removed from the list of differentially expressed genes. Finally the genes list was annotated including gene names, GeneBank accession code, description and gene ontology. For some of the probes this information was not available and they conserved the Affymetrix probe code.

CELL COUNTING

The fluorescent *in situ* hybridisation Z-stacks images of the embryonic trunk region used for counting cells were acquired using the Zeiss Axio Imager.M2 microscope with the Apotome technology. To count the cells we used a cell counter plug-in in the ImageJ software that allows you to mark each cell you have counted with the mouse.

It also keeps those markers in the different planes of the Z-stack to avoid double counting of the same cell present in adjacent planes. Once you have gone through all the planes in a stack, the plug-in gives you the number of cells you have counted.

IMAGE PROCESSING

Microscope images were processed using the Zeiss's microscopy software ZEN and Adobe Photoshop CS6 to cut the region of interest of our photographs and adjust the contrast and brightness. The SciFig plug-in for ImageJ was used to create the image panels.

Chapter 3:

Expanding the GRN

Introduction

Greenhill *et al.* (2011) set up the core GRN underlying melanocyte specification and differentiation *in vivo* through a systems biology approach. This GRN includes some of the main genes involved in this process, but there are still many other known and unknown genes that need to be incorporated in the model to understand the basis for stable melanocyte differentiation. Several methods can be used to identify new players involved in the melanocyte GRN. The classic method is to perform *in situ* hybridisation using a mutant of a gene affecting melanocyte pigmentation. Then we evaluate the gene expression of other genes that may be related to melanocyte biology. A reduction or loss of expression in the *in situ* hybridisation indicates that the gene is regulated by the original mutant gene, and hence it would be involved in melanocyte development. The limitations with this method are that we need to have some prior knowledge of the candidate genes to test (and produce the ISH probes), and we do not know if the interaction is direct or not.

Another method that can be used and is also based on melanocyte mutant fish is microarrays. This is a high throughput method, which means that we can test thousands of genes at a time, but unfortunately in microarray experiment a probe is needed for every gene in the genome, which implies that a prior knowledge of the organism's genome is needed. The method is based on performing a microarray experiment comparing gene expression in melanocyte mutant fish and wild type fish. The genes that are up or down-regulated in the mutant would be predicted to be regulated by the original mutated gene. However, this method presents some limitations: we do not know if the regulation is direct or not, and also it may not be very sensitive if we use the whole organism as some genes may be highly expressed in tissues that we are not studying, masking the important results.

An evolution of this method is based on RNAseq. This new technology uses the capabilities of high throughput sequencing to identify all the RNA molecules present in a cell or tissue at a certain time, hence it determines all the genes expressed at that stage. This can be used to compare the transcriptome of a melanocyte mutant with a wild type fish, following the same logic as in the microarray experiment (Wang *et al.*, 2009). It has the same advantages and issues of the microarray-based methods.

Lastly, if we want to determine if a certain transcription factor directly regulates a target gene we can use chromatin immunoprecipitation (ChIP) integrated with sequencing (ChIP-seq) or microarray hybridisation (ChIP-chip). ChIP is based on the in vivo cross-linking of the chromatin at the moment in time we want to study, then the chromatin is extracted and fragmented, finally an antibody against the TF we are studying is used to immunoprecipitate the DNA fragments bound to the TF. The cross-linking is then reversed and this is followed by a sequencing step to identify the bound DNA fragments, or the hybridisation to a DNA array to find the gene bound to the TF (Carey *et al.*, 2009). The approach chosen in this project was to use microarray analysis to identify candidate genes for the melanocyte GRN.

Chipperfield (2009) used a microarray approach to identify candidate direct and indirect targets of Sox10 by comparing the expression profiles of *sox10* expressing cells from both wild type and *sox10* mutant embryos. Those genes whose expression is regulated by *sox10* will appear downregulated in *sox10* mutants. The transgenic lines Tg(-7.2*sox10*:EGFP) and Tg(-4.9*sox10*:EGFP), that express GFP under the control of a 7.2 kb and 4.9 kb region of the *sox10* promoter respectively, were used for the microarray experiments. These fish display GFP expression in the neural crest derived cells although the Tg(-7.2*sox10*:EGFP) line shows a stronger GFP expression in the otic epithelium than the other line (Carney *et al.*, 2006; Dutton *et al.*, 2008).

GFP expressing cells were purified by fluorescent activated cell sorting (FACS). Gene expression levels are compared between GFP positive cells from WT embryos carrying the transgenic lines and GFP positive cells from *sox10* mutant embryos using the Affymetrix zebrafish GeneChip microarray, which can detect almost 15000 different transcripts (Affymetrix, 2008). In this system, each independent sample is hybridised to an array and the comparisons are made between arrays (in this case

WT versus mutant arrays). Microarray data was generated using 24 hpf embryos as this is when neural crest cells are undergoing and expressing key genes for specification, differentiation and migration, as well as being the earliest stage at which mutant and WT embryos can be distinguished. The data from the Tg(-7.2*sox10*:EGFP) line consisted of three WT and three mutant arrays, and the data from the Tg(-4.9*sox10*:EGFP) consisted of two WT and two mutant arrays.

Chipperfield performed the microarray analysis using the freely available web tool GEPAS (Gene Expression Profile Analysis Suite, www.gepas.org). To normalise the data, the expresso module on the GEPAS website was used. The analysis of differentially expressed genes was carried out using the T-rex tool from GEPAS with the T-test option. This generated a list of genes ranked by t-statistic with downregulated genes presenting a negative value and upregulated genes positive values. No correction methods for multiple testing were applied in the analysis.

From the top 100 most downregulated genes in *sox10* mutant (Table 5), 89 were analysed through ISH for validation. 25 of them were validated as downregulated in *sox10* mutants. Some of these genes that were expressed in the melanocyte lineage were already well known melanocyte markers such as *mitfa* or *dct*. 5 were novel genes that were previously unknown to have a role in melanocyte biology and to be regulated by *sox10*: *atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36* (Figure 40 Appendix 2).

The process to generate *sox10* mutants expressing GFP under the control of the *sox10* promoter is fully explained in Chipperfield (2009) thesis, as well as all the details of the FACS, microarray and *in situ* hybridisation experiment.

Table 5. List of the 100 most downregulated genes in *sox10* mutant embryos, ranked by t-statistic from Chipperfield's thesis. The most up to date gene names are provided in brackets. The list shows the 100 most down-regulated genes identified by Chipperfield from the 7.2*sox10*:GFP microarray data analysis using the GEPAS online tool. The genes are ranked by t-statistic that was obtained from the T-test analysis.

Gene Symbol	t-statistic
Dr.16789.1.A1_at	-25.393
wu:fc31e04	-18.127
Dr.13919.1.S1_at (fzd7b)	-17.892
LOC565706 (pde4ba)	-16.778
LOC564804 (atp6v1ab)	-16.267
cldnj	-16.124
rbp4l	-16.037
sb:cb319	-15.54
wu:fj94a03 (rbm47)	-14.789
bhlhb3l	-14.48
ndrg1	-12.15
zgc:154036 (wwp2)	-12.091
ift88	-11.918
pik3c3	-11.487
Dr.25879.1.A1_at	-11.355
wu:fc46b01, zgc:100933	-11.345
nfkb2	-10.857
zgc:158673 (si:dkey-23o10.3)	-10.73
tmem169	-10.723
Dr.22614.1.A1_at (myo5b)	-10.704
pax7	-10.592
wu:fj38e08	-10.548
zgc:110397 (tada2b)	-10.191
slc2a15b	-9.988
ccbl2	-9.97
fabp7a	-9.913
bhlhb2 (bhlhe40)	-9.713
wu:fc17h11	-9.55
ell	-9.492
mg:cb01g09 (mfn2)	-9.401
coro1c	-9.323
wu:fi46h04	-8.871
wu:fc54b10 (mon1bb)	-8.801
zgc:110718 (ankrd49)	-8.793
dohh	-8.766
si:dkeyp-86b9.2 (vamp5)	-8.714
LOC568088 (rbm25a)	-8.689
cx33.8 (cx30.3)	-8.688
zgc:112072 (slc24a4a)	-8.672
atic	-8.667
wu:fi05a09 (gmps)	-8.556
Dr.25954.1.A1_at	-8.543
foxd3	-8.529

ppa1	-8.385
Dr.3972.1.S1_at (si:dkey-226k3.4)	-8.302
zgc:110343 (prdx1)	-8.217
keap1	-8.161
mitfa	-8.03
zgc:152987 (si:ch211-137a8.2)	-7.965
wu:fb95f11	-7.965
wu:fc08b04 (dync1li2)	-7.946
zgc:100919 (tspan36)	-7.945
znf593	-7.898
cct5	-7.89
wu:fj82h06	-7.88
pi4k2a	-7.83
Dr.26101.1.A1_at	-7.819
paics	-7.758
aldh2 (aldh2.1/aldh2.2)	-7.748
crestin	-7.728
Dr.18158.1.A1_at	-7.71
zgc:110343 (prdx1)	-7.705
wu:fa98e08	-7.673
wu:fj68b05	-7.478
pah	-7.446
slc2a15b	-7.372
Dr.14845.1.S1_at	-7.359
Dr.13642.1.S1_at	-7.181
zgc:66482 (pik3ip1)	-7.173
mkks	-7.126
adsl	-7.087
derl1	-7.018
dct	-6.987
cyp3c1l2	-6.98
Dr.4615.2.S1_at	-6.975
mtch2	-6.947
aldh2b	-6.883
LOC100004591	-6.873
zgc:112247 (sptlc1)	-6.854
zgc:162119 (pla2g12a)	-6.705
wu:fd02h10	-6.692
zgc:110239	-6.676
atp6v1e1	-6.639
otomp	-6.638
si:dkey-180p18.9	-6.633
zgc:63767 (slc15a4)	-6.522
Dr.21407.1.A1_at	-6.482

zgc:65870 (pcbp4)	-6.428
cotl1	-6.401
wu:fb74h01	-6.4
si:ch211-243g18.2	-6.365
s100b	-6.307
si:ch211-218c6.1 (zc3h18)	-6.296
Dr.10151.1.A1_a_at (capn5a)	-6.282
Dr.7955.1.A1_at	-6.274
degs1	-6.27
phf8	-6.231
slc25a26	-6.227
zgc:113142	-6.185
ggps1	-6.173

We sought to use the microarray data obtained by Chipperfield (2009) to examine it using the most up to date microarray analysis algorithms and the most recent gene expression information available in the ZFIN database to discover new candidate genes with an important function in melanocyte development that will be incorporated to the melanocyte gene regulatory network.

Results

Microarray analysis using R

Chipperfield performed this analysis several years ago and the algorithms and tools to analyse microarrays have advanced in this time. Also, coming from a bioinformatics background I have some experience in microarray analysis, so I decided to get his raw data and repeat his analysis to verify whether I would obtain similar results with the up to date algorithms.

The analysis of the Affymetrix Cell Intensity Files (.cel files) that Chipperfield (2009) produced was performed using the Bioconductor package for the R statistical programming language. The objective was to analyse microarray data using a more versatile and better established approach in the bioinformatics community than the one he used before, which was based on a website that is no longer maintained and available, asking whether I obtained a similar list of differentially expressed genes.

An R script (Appendix 3) was produced where the ".cel" files were loaded and then the data was pre-processed to eliminate the variation between arrays due to non-biological factors such as sample processing. To get the differentially expressed genes the T-Test statistical test was used. The T-Test produced a list of genes ranked by t-statistics with the most downregulated genes at the top of the list (with most negative values). The T-Test produced an associated p-value for each gene, which measures the probability of the test statistic being equal or greater than the observed result assuming that the null hypothesis is true. If the p-value is smaller or equal to the significance level then the null hypothesis (no difference between wild type and mutant embryos) can be rejected (Chipperfield, 2009). In the T-Test all the genes with a p-value lower or equal than 0.05 were selected. Finally the gene list was annotated with the most up to date gene names and a table was produced (Table 6). This analysis was carried out using the same settings that Chipperfield used in 2009 so the list could be compared.

Table 6. List of 100 top downregulated genes from my analysis using the Bioconductor package for R and the same settings that Chipperfield used in 2009. The list shows the 100 most down-regulated genes from the 7.2*sox10*:GFP microarray data analysis ranked by t-statistics. The T-Test produces an associated p-value for each gene, which measures how significantly downregulated each gene is. These genes have been selected for having a p-value lower or equal than 0.05, making them significantly downregulated.

Gene Symbol	**t-statistics**
Dr.16789.1.A1_at	-25.368
wu.fc31e04	-18.128
fzd7b	-17.892
pde4ba	-16.776
atp6v1ab	-16.254
cldnj	-16.125
rbp4l	-16.037
sb.cb319	-15.541
rbm47	-14.803
bhlhe41	-14.480
ndrg1a	-12.161
wwp2	-12.063
ift88	-11.918
pik3c3	-11.486
Dr.25879.1.A1_at	-11.359
wu:fc46b01	-11.345
nfkb2	-10.848

tmem169	-10.734
si:dkey-23o10.3	-10.728
myo5b	-10.704
pax7a	-10.592
wu.fj38e08	-10.550
tada2b	-10.191
slc2a15b	-9.988
ccbl2	-9.977
fabp7a	-9.913
bhlhe40	-9.715
wu.fc17h11	-9.533
ell	-9.495
mfn2	-9.431
coro1ca	-9.323
wu:fi46h04	-8.865
mon1bb	-8.804
ankrd49	-8.796
dohh	-8.742
vamp5	-8.714
cx30.3	-8.689
rbm25a	-8.688
slc24a4a	-8.672
atic	-8.667
gmps	-8.553
Dr.25954.1.A1_at	-8.541
foxd3	-8.528
ppa1b	-8.423
wu:fc84a10	-8.302
prdx1	-8.217
keap1a	-8.166
mitfa	-8.030
cct5	-7.975
si:ch211.137a8.2	-7.966
wu:fb95f11	-7.961
dync1li2	-7.948
tspan36	-7.945
znf593	-7.901
wu.fj82h06	-7.880
pi4k2a	-7.828
Dr.26101.1.A1_at	-7.820
paics	-7.762
aldh2.1/aldh2.2	-7.750
crestin	-7.731
prdx1.1	-7.705

Dr.18158.1.A1_at	-7.705
wu.fa98e08	-7.673
wu.fj68b05	-7.477
pah	-7.447
slc2a15b.1	-7.372
Dr.14845.1.S1_at	-7.357
Dr.13642.1.S1_at	-7.181
pik3ip1	-7.174
mkks	-7.126
adsl	-7.087
derl1	-7.032
dct	-6.987
cyp3c1l2	-6.979
Dr.4615.2.S1_at	-6.967
mtch2	-6.949
aldh2.2	-6.882
LOC100004591	-6.873
sptlc1	-6.865
pla2g12a	-6.703
wu.fd02h10	-6.691
zgc.110239	-6.670
atp6v1e1b	-6.638
otomp	-6.638
si.dkey.180p18.9	-6.632
slc15a4	-6.522
Dr.21407.1.A1_at	-6.482
pcbp4	-6.428
wu.fb74h01	-6.403
cotl1	-6.399
si.ch211.243g18.2	-6.365
s100b	-6.307
zc3h18	-6.301
capn5a	-6.283
Dr.7955.1.A1_at	-6.274
degs1	-6.270
phf8	-6.231
slc25a26	-6.230
zgc.113142	-6.185
ggps1	-6.173

Reassuringly, the list obtained by this analysis was the same as the one obtained by Chipperfield in 2009. Some of the gene names in the Chipperfield list were out of date, as it was produced over 6 years ago. Hence, his list was updated with the new

gene names found in the ZFIN database, which allowed us to establish that both lists contain the same genes, with the same p-values.

Sox10 dependent genes are also Mitfa dependent

Chipperfield performed an ISH screen to validate the genes identified as downregulated in *sox10* mutant embryos in the microarray analysis. He synthesised probes and performed ISH for 89 genes using wild type and *sox10* mutant embryos at 30 hpf. From these genes many were false positives, and some others were validated as expressed in the xanthophore and otic vesicle. This study validated 5 novel genes as downregulated in 30 hpf *sox10* mutant embryos in the melanocyte lineage. However, this study left open the question of whether these genes might be regulated by some other TF downstream of *sox10*. In order to find out if these genes were directly or indirectly regulated by *mitfa*, ISH was done using 30hpf *mitfa* mutant embryos. The same probes synthesised by Chipperfield were used in this experiments. Figure 7 shows that all the 5 genes are expressed in melanocytes in wild type embryos, and are downregulated in *mitfa* mutant embryos. This indicates that all are downstream of Mitfa in the melanocyte GRN.

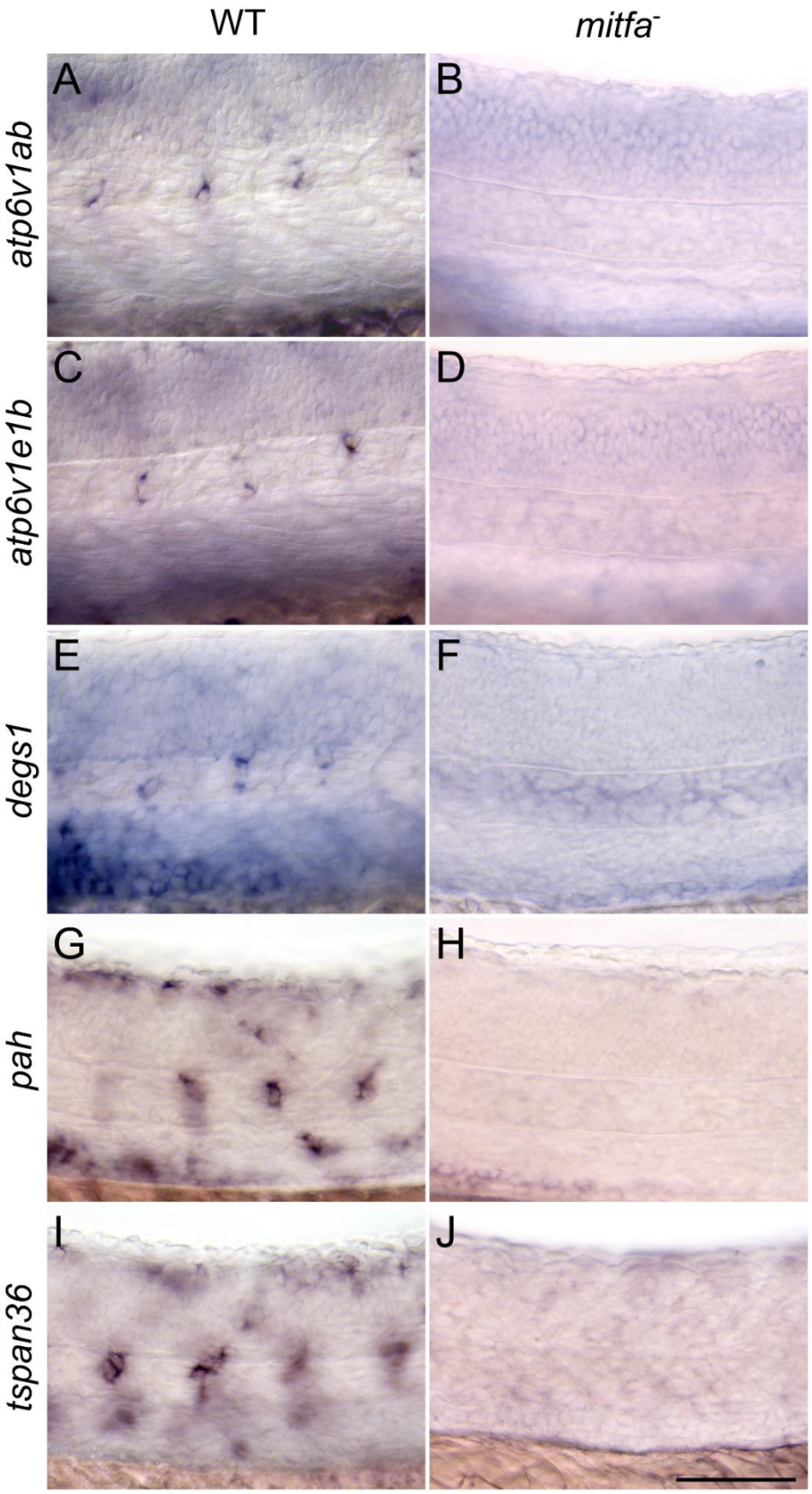

Figure 7. ISH expression pattern of the 5 verified melanocyte genes in 30 hpf wild type and *mitfa* mutant embryos. All images are lateral views of the mid trunk. All 5 genes (*atp6v1ab*, *atp6v1e1b*, *degs1*, *pah* and *tspan36*) are expressed in melanocytes of wild type embryos but

are not expressed in *mitfa* mutant embryos. Image orientation: embryo head is on the left and tail on the right of the images. Scale bar: 100 µm

Analysis of the combined microarray dataset using improved script

Chipperfield performed the same process of FACS sorting and microarray hybridisation with the transgenic line Tg(-4.9*sox10*:EGFP) a few months after he did the 7.2*sox10*:GFP line. These two lines display similar GFP expression, but the 4.9 line displays a weaker otic epithelium expression. However for our study this is not very important as we want to identify genes involved in pigment cell development. The data from the 4.9 line consisted of two WT versus two mutant arrays. Due to time constraints Chipperfield only used the data from the 7.2*sox10*:EGFP microarray analysis to perform the ISH screen to validate the *sox10* targets, as he gathered the 7.2*sox10*:GFP data before the 4.9*sox10*:EGFP. The aim of this section was to use the combined dataset to perform a more robust statistical analysis of the differentially expressed genes.

All the 10 microarrays, 6 from the 7.2 line and 4 from the 4.9 line were used in the analysis. For the pre-processing we used the threestep function with the default settings: RMA background correction, quantile normalisation, no perfect match evaluation and median polish summarisation (Gautier *et al.*, 2004). Then we performed a filtering step to remove the genes that are absent in all samples. In order to identify the differentially expressed genes between mutant and wild type embryos the eBayes statistical test is applied (Ritchie *et al.*, 2015). This test produces more reliable results than T-test with small samples. The test calculates an associated p-value that measures the significance of a given gene to be downregulated. The p-values were corrected using the Benjamini and Hochbergh method to control the false positive rate (Pollard *et al.*, 2004). These adjusted p-values were used to rank the genes, where the lower the p-value the higher the significance of that gene to be downregulated in mutant versus wild type embryos. A significance level of 0.05 was used, and all the genes with an adjusted p-value higher than that were excluded.

Finally, the genes were annotated with their gene symbol and they were saved into a table (Table 7). The resulting gene list contained 113 genes.

Table 7. List of the 113 downregulated genes obtained from the combined dataset analysis with a p-value lower than 0.05 after applying the FDR adjustment. The genes highlighted in green were also present in the list obtained in the analysis where we used just the 7.2 arrays dataset and the t-test statistical analysis (Table 6).

Gene Symbol	Adjusted p-value
bhlhe41	0.0000
rbp4l	0.0001
wu:fc46b01	0.0001
si:dkey.226k3.4 (wu:fc84a10)	0.0001
npr3	0.0004
fzd7b	0.0004
wu.fc31e04	0.0004
sb.cb319	0.0004
atp6v1ab	0.0007
slc2a15b	0.0008
pik3c3	0.0008
hoxc8a	0.0009
pah	0.0009
si.dkey.180p18.9	0.0010
tspan36	0.0010
dct	0.0012
ca2	0.0012
prdx1.1	0.0012
aldh2.1/aldh2.2	0.0013
atp6v1e1b	0.0015
ndrg1a	0.0015
s100b	0.0025
aldh2.2	0.0031
Dr.26101.1.A1_at	0.0031
pvalb7	0.0033
Dr.16789.1.A1_at	0.0035
pik3ip1	0.0035
ginm1	0.0037
atic	0.0043
paics	0.0054
slc2a15b.1	0.0056
prdx1.2	0.0057
Dr.4838.1.A1_at	0.0059
slc24a4a	0.0073
prdx1	0.0084
si.dkey.73n10.1	0.0084
nradd	0.0098
Dr.10941.1.A1_at	0.0098
vac14	0.0098
uap1	0.0098

mitfa	0.0104
atp6v1g1	0.0105
fbp1a	0.0117
rab32a	0.0133
bace2	0.0133
vamp5	0.0133
gmps	0.0133
coro1ca	0.0139
sulf2a	0.0150
keap1a	0.0150
ap5m1	0.0159
jag2	0.0159
dmtn	0.0160
pcbp4	0.0166
adsl	0.0172
wu.fj65h10	0.0172
vegfc	0.0184
syngr1a	0.0185
wu.fc26d05	0.0186
ednrba	0.0187
mon1bb	0.0187
wu.fd02h10	0.0203
zic2b	0.0204
Dr.20434.1.A1_at	0.0204
dlb	0.0239
si.ch211.195b13.1	0.0240
fkbp1ab	0.0246
ankrd27	0.0261
Dr.752.1.A1_at	0.0261
pcdh10a	0.0269
cdon	0.0269
dcn	0.0269
ppp1r13ba	0.0269
lrrtm4l1	0.0269
slc37a2	0.0269
pla2g15	0.0269
gart	0.0269
zgc:154045	0.0269
slc3a2a	0.0269
pank1a	0.0269
rnaseka	0.0269
wu.fj94h02	0.0269
sigmar1	0.0277
mov10b.1	0.0292
atp6ap1b	0.0296
Dr.14210.1.A1_at	0.0306
acss1	0.0318
epha6	0.0322
atp6v1f	0.0322
vps26a	0.0322

ppat	0.0322
bhlhe40	0.0322
cx30.3	0.0322
wu.fj82h06	0.0331
myo5b	0.0336
crestin	0.0338
pttg1ipb	0.0368
sb.cb8	0.0383
rab7	0.0397
hoxc8a.1	0.0413
cyb5a	0.0427
tspan18a	0.0432
wu.fj35f12	0.0432
rcan2	0.0432
fa2h	0.0432
tnfrsf9a	0.0432
wu.fa98e08	0.0432
hapln1b	0.0432
jagn1b	0.0435
cd63	0.0456
slco3a1	0.0459
sypl2b	0.0459
fabp7a	0.0476

45 out of the 113 genes (highlighted in green) were also present in the top 100 downregulated genes obtained using the 7.2 microarrays and the t-test analysis (Table 6). This analysis has found four of the genes that were validated by ISH as downregulated in *sox10* mutants in the melanocyte lineage using the previous analysis: *atp6v1ab, atp6v1e1b, pah* and *tspan36* (Figure 7). Degs1 was not found in this analysis despite having been validated as driven by *sox10*.

ZFIN expression pattern of novel candidate genes

The analysis of the combined microarray dataset has provided 68 new candidate genes that might be downregulated in *sox10* mutants. These genes should be validated using an ISH screen approach similar to the one used by Chipperfield in 2009, unfortunately due to time constraints it was not possible to perform the ISH with all these genes. The approach that we selected was to use the online database ZFIN (Howe *et al.*, 2013) to find out the published expression pattern images and the annotated expression structures of these genes to check if they are expressed in the

neural crest or pigment cells. This was done by manually searching for every gene on the database, then identifying the expression structures that have been annotated, and examining the expression pattern images where available. All these data has been collated in Table 8, with a description of the expression pattern image.

Table 8. Expression patterns of the 68 new candidate genes with their ZFIN expression structures tags and a description of the expression pattern images when available. Most of the genes are expressed in many tissues in the fish and therefore have been annotated with many expression structure tags. Here we include the genes annotated with the neural crest or pigment cell tags and use the word "others" if they are expressed in other structures. A subjective description of the expression pattern image is also included when available. The genes highlighted in yellow are the ones annotated with the pigment cell tag. The genes highlighted in blue are the ones annotated with the neural crest tag or those whose expression pattern image suggests a neural crest or pigment cell expression.

Gene Symbol	**ZFIN Expression structures**	**Pattern description**
npr3	others	No neural crest or pigment cells expression
hoxc8a	others	It's widely expressed. Difficult to reject neural crest or pigment cells expression.
ca2	others	No neural crest or pigment cells expression
pvalb7	others	No neural crest or pigment cells expression
ginm1	others	Images aren't detailed enough. Difficult to tell.
Dr.4838.1.A1_at	No information	
si:dkey-73n10.1	No information	
nradd	No information	
Dr.10941.1.A1_at	No information	
vac14	others	Images aren't detailed enough. Difficult to tell.
uap1	others	Images show diffused expression along the trunk. Difficult to tell.
atp6v1g1	others	Images show expression along the trunk. Difficult to tell.
fbp1a	others	No neural crest or pigment cells expression
rab32a	neural crest, pigment cell, trunk neural crest cell and others	Very likely pigment cells expression. Worth exploring further.
bace2	neural crest, pigment cell and others	Very likely pigment cells expression. Worth exploring further.
sulf2a	others	No neural crest or pigment cells expression
ap5m1	No information	
jag2	others	No neural crest or pigment cells expression
dmtn	others	No neural crest or pigment cells expression
wu:fj65h10	No information	

vegfc	others	No neural crest or pigment cells expression
syngr1a	neural crest, pigment cell and others	Very likely pigment cells expression. Worth exploring further.
wu:fc26d05	others	
ednrba	head neural crest, iridophore, pigment cell, trunk pigment cell	Known to be expressed in pigment cells.
zic2b	cranial neural crest and others	No neural crest or pigment cells expression
Dr.20434.1.A1_at	No information	
dlb	others	No neural crest or pigment cells expression
si:ch211-195b13.1	No information	
fkbp1ab	others	No neural crest or pigment cells expression
ankrd27	No information	
Dr.752.1.A1_at	No information	
pcdh10a	neural crest and others	Expressed in migrating neural crest. Further study needed.
cdon	neural crest and others	No neural crest or pigment cells expression
dcn	others	No neural crest or pigment cells expression
ppp1r13ba	others	No neural crest or pigment cells expression
lrrtm4l1	others	No neural crest or pigment cells expression
slc37a2	No information	
pla2g15	No information	
gart	neural crest, pigment cell and others	Very likely pigment cells expression. Worth exploring further.
zgc:154045	No information	
slc3a2a	others	Shows a spotty pattern, so it may be expressed in pigment cells. Further study needed.
pank1a	others	No neural crest or pigment cells expression
rnaseka	No information	
wu:fj94h02	No information	
sigmar1	others	No neural crest or pigment cells expression
mov10b.1	No information	
atp6ap1b	pigment cell and others	Very likely pigment cells expression. Worth exploring further.
Dr.14210.1.A1_at	No information	
acss1	others	No neural crest or pigment cells expression
epha6	No information	
atp6v1f	others	No neural crest or pigment cells expression
vps26a	others	No neural crest or pigment cells expression
ppat	neural crest and others	Image is not very clear but it may be expressed in pigment cells. Further study needed.

pttg1ipb	neural crest, pigment cell and others	Very likely pigment cells expression. Worth exploring further.
sb:cb8	others	No neural crest or pigment cells expression
rab7	neural crest and others	Pattern suggest that it may be expressed in pigment cells. Further study needed.
hoxc8a.1	others	No neural crest or pigment cells expression
cyb5a	neural crest and others	Shows a spotty pattern, so it may be expressed in pigment cells. Further study needed.
tspan18a	No information	
wu:fj35f12	No information	
rcan2	No information	
fa2h	others	No neural crest or pigment cells expression
tnfrsf9a	No information	
hapln1b	others	No neural crest or pigment cells expression
jagn1b	No information	
cd63	others	No neural crest or pigment cells expression
slco3a1	others	No neural crest or pigment cells expression
sypl2b	neural crest and others	Image is very faint but it suggest that it may be expressed in pigment cells. Further study needed.

From the ZFIN database search (Table 8) there are 7 genes (highlighted in yellow in the table) whose expression pattern is very similar to the one of other pigment cell markers and they are annotated as expressed in pigment cells: *rab32a, bace2, syngr1a, ednrba, gart, atp6ap1b* and *pttg1ipb*. The pigment pattern available in ZFIN is from Thisse *et al.* 2001, 2004 and 2005, and the images of the ISH carried out using wild type fish in stages between Prim 15 (30 hpf) and Prim 25 (36 hpf) are shown in Figure 8 and Figure 9 for reference.

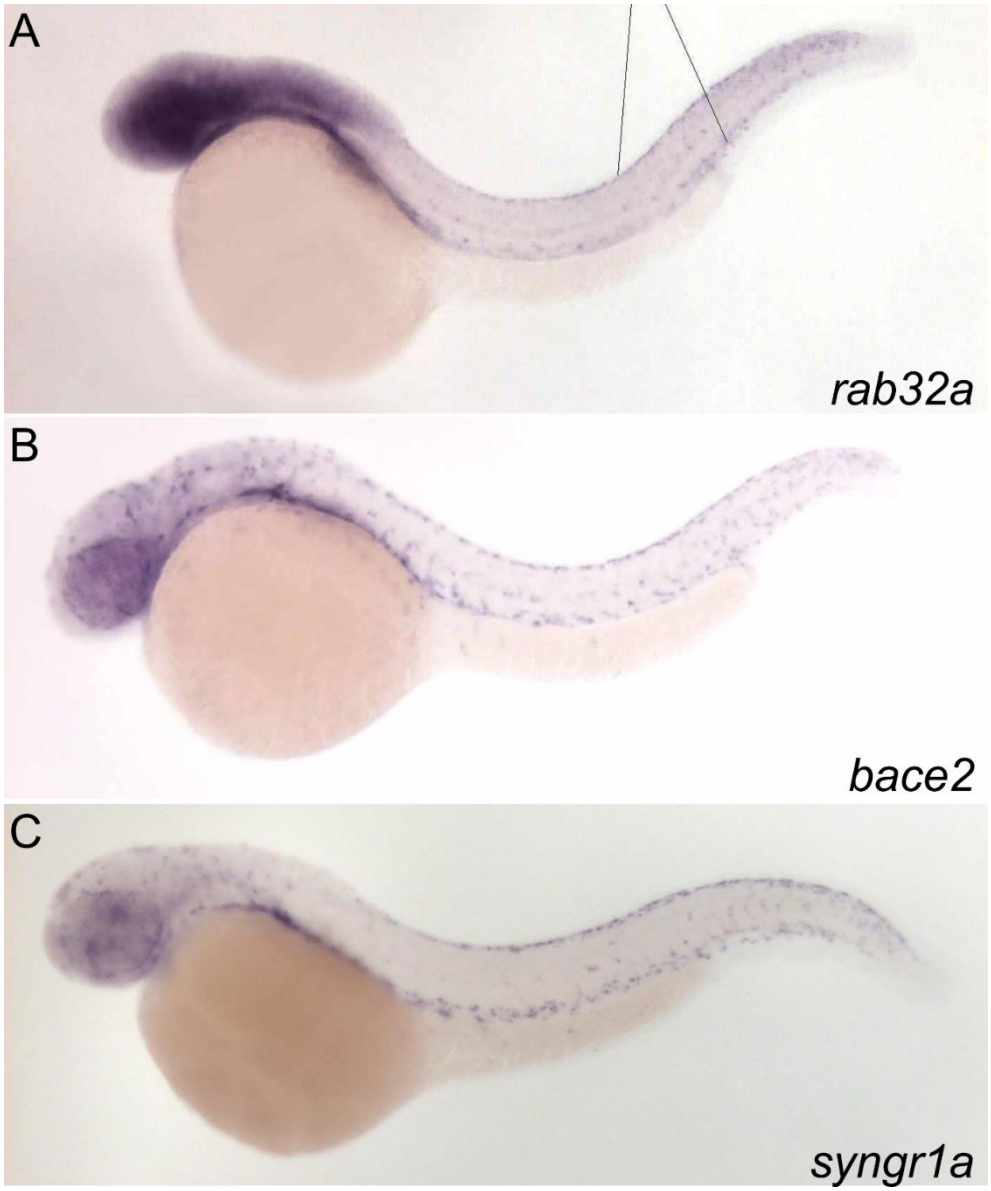

Figure 8. ISH expression pattern of *rab32a* (A), *bace2* (B) and *syngr1a* (C) genes from wild type embryos. The images are from the ZFIN database and were submitted by Thisse *et al.* 2001, 2004 and 2005. The ISH are from stages in between Prim 15 (30 hpf) and Prim 25 (36 hpf). Image orientation: embryo head is on the left and tail on the right of the images.

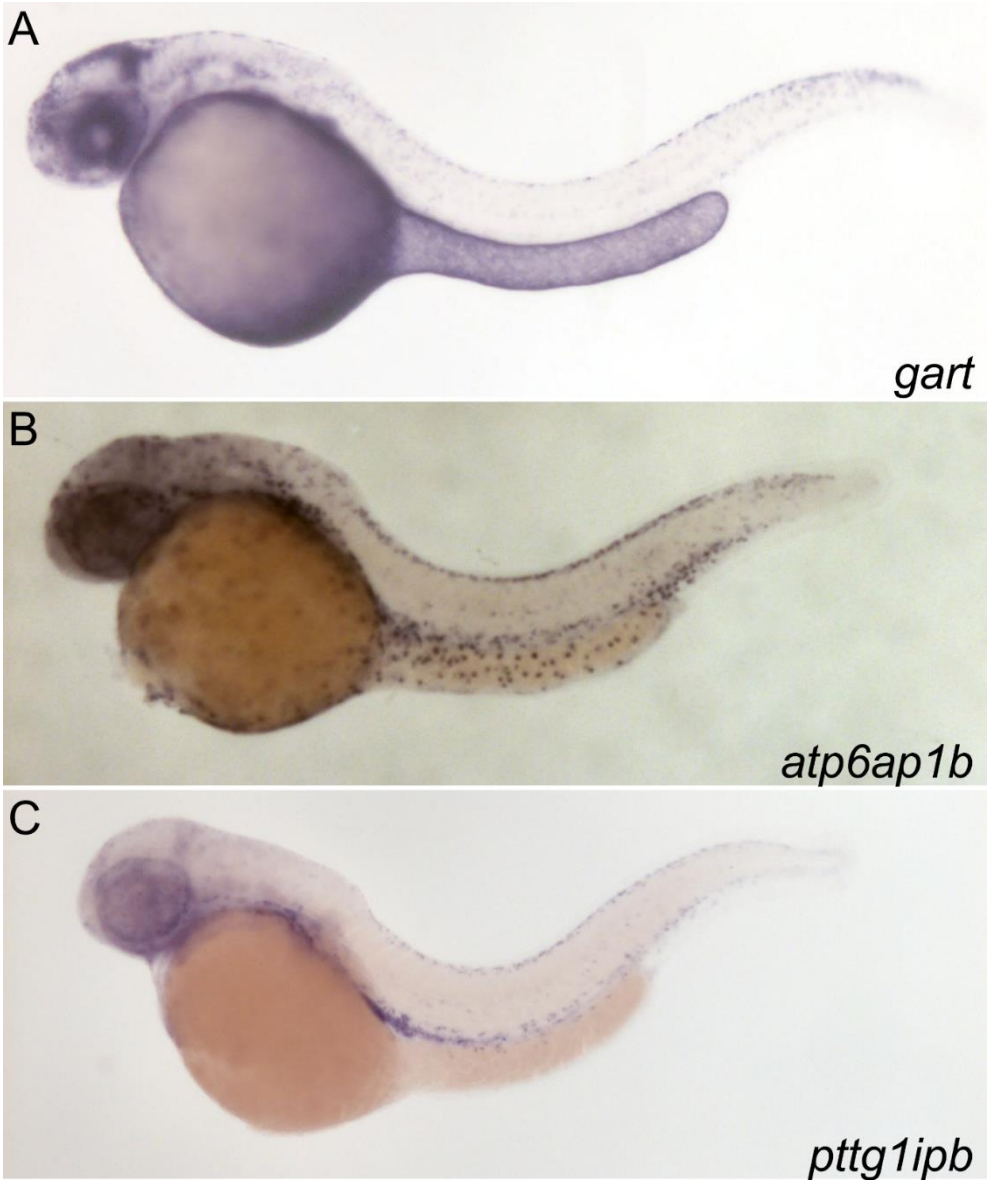

Figure 9. ISH expression pattern of *gart* (A), *atp6ap1b* (B) and *pttg1ipb* (C) genes from wild type embryos. The images are from the ZFIN database and were submitted by Thisse *et al.* 2001, 2004 and 2005. The ISH are from stages in between Prim 15 (30 hpf) and Prim 25 (36 hpf). Image orientation: embryo head is on the left and tail on the right of the images.

The quality of the ISH images submitted to ZFIN for these genes is not perfect, but it is good enough to observe that most of these genes are expressed in the dorsal and ventral stripe of the trunk, as well as in cells migrating on the medial pathway, locations that correspond with the places where we expect to find pigment cells. Ednrba (also highlighted in yellow in Table 8) had already been reported as expressed in the pigment cell lineage by Parichy *et al.* (2000).

There are some genes (highlighted in blue) in the Table 8 list whose expression pattern is not very clear in the images, but which were suggestive of expression in neural crest or pigment cells: *pcdh10a, slc3a2a, ppat, rab7, cyb5a* and *sypl2b*. Their expression patterns from wild type fish in stages from Prim 15 (30 hpf) to Prim 25 (36 hpf) obtained from the ZFIN database (Thisse *et al.* 2001, 2004 and 2005) are shown in Figure 10 for reference. As the expression pattern is not as clear as in the previous 7 genes, they will need to be studied more in depth to determine if they are involved in melanocyte development.

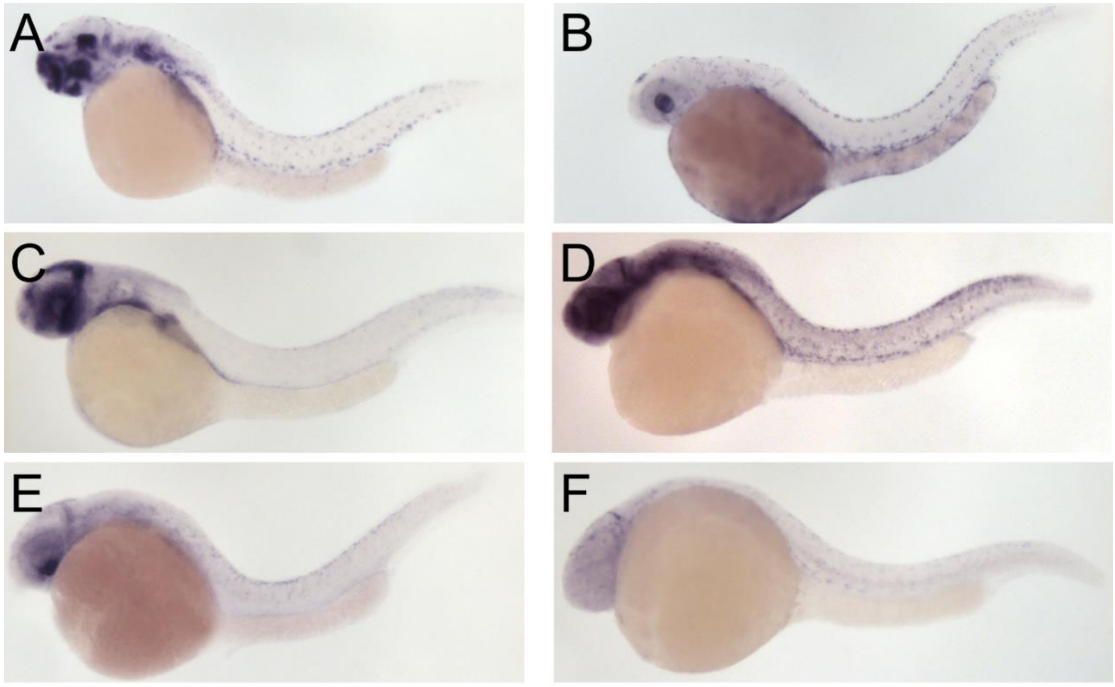

Figure 10. ISH expression pattern of *pcdh10a* (A), *slc3a2a* (B), *ppat* (C), *rab7* (D), *cyb5a* (E) and *sypl2b* (F) genes from wild type embryos. The images are from the ZFIN database and were submitted by Thisse *et al.* 2001, 2004 and 2005. The ISH from stages in between Prim 15 (30 hpf) and Prim 25 (36 hpf). Image orientation: embryo head is on the left and tail on the right of the images.

Location of Sox10 or Mitfa binding sites in candidate genes

All of the previous 13 genes were found to be downregulated in *sox10* mutant embryos in the microarray analysis, therefore they are regulated (directly or not) by Sox10 transcription factor. Some of these genes may be part of the melanocyte GRN, but they might also have a role in other pigment cells biology (iridophore or xanthophore) or other *sox10* expressing cells (glia, oligodendrocytes or sensory neurons). An analysis of their function from the published literature may help us understand if these genes have a function in melanocyte biology or in other cell types. If they have a melanocyte function, it will be helpful to know it they are directly regulated by *mitfa* (the melanocyte master regulator), as this would indicate that they are part of the melanocyte GRN and will help us to find its location within the network. That is why we decided to carry out an *in silico* search of Mitfa binding sites in these gene's promoters.

The first 7 genes were annotated as being expressed in pigment cells and their ZFIN expression patterns also suggests that. Therefore these were strong candidates to investigate if they were part of the melanocyte GRN. In order to determine their position in the network we need to know which gene regulates them. As they have been obtained from a microarray analysis comparing *sox10* mutants with WT embryos, we are going to test if they are directly or indirectly regulated by *sox10* or *mitfa* (as a transcription factor gene that is a direct downstream target of *sox10 a*nd a master regulator of melanocyte development). For the other 6 genes the expression pattern is a bit unclear, but they may also be part of the network so it may be interesting to know if they are regulated by Sox10 or Mitfa.

In order to find out if these 13 genes are regulated by Sox10 or Mitfa, a program was designed to find their binding site in their promoter. A region of 3000 bps of the promoter and 5'UTR (Kristiansson *et al.*, 2009) was downloaded from the Ensembl genome repository (*Danio rerio* GRCz10 version) using the Biomart tool in Ensemble. Then a Perl program was designed to look for the binding site motifs of the Sox10 or Mitfa transcription factors in these gene's promoters (Appendix 4).

The study carried out by Elworthy *et al.* (2003) demonstrated that Sox10 directly regulates the expression of *mitfa*. Mitfa promoter has two Sox10 binding sites in melanocytes in the zebrafish embryo. The site S1 (CTCAAAG) is necessary for *mitfa* expression; the site S3 (GATTGTA) also drives *mitfa* expression but to a lesser extent.

Regarding Mitfa binding sites, no sequence has been identified yet in zebrafish, however Pogenberg *et al.* (2012) stablished that in mouse MITF has the ability to bind to the palindromic CACGTG E-box motif. Moreover, unlike other bHLHZip transcription factors, MITF also binds to the asymmetric TCATGTG M-box sequence.

The *mitfa* and *dct* promoters were used as a positive control, as they are known to be directly regulated by Sox10 and Mitfa respectively (Elworthy *et al.*, 2003; Ludwig *et al.*, 2004). Two genes were used as negative control: *cherp* and *mif*. These genes were randomly selected from the list of differentially expressed genes between mutant and wild type embryos as ones whose gene expression was not significantly different i.e. their p-value is close to 1. The expression of these genes is not supposed to be controlled by Sox10, and hence the Perl program should not find Sox10 or Mitfa binding sites on their promoters. This analysis tests if the program finds random Sox10 or Mitfa binding sites in promoters of genes that are not related to pigment cell biology.

The Perl program takes the promoters of the 17 genes as input and looks for the Sox10 or Mitfa binding sites, telling us the genes that contain these motifs in their promoter.

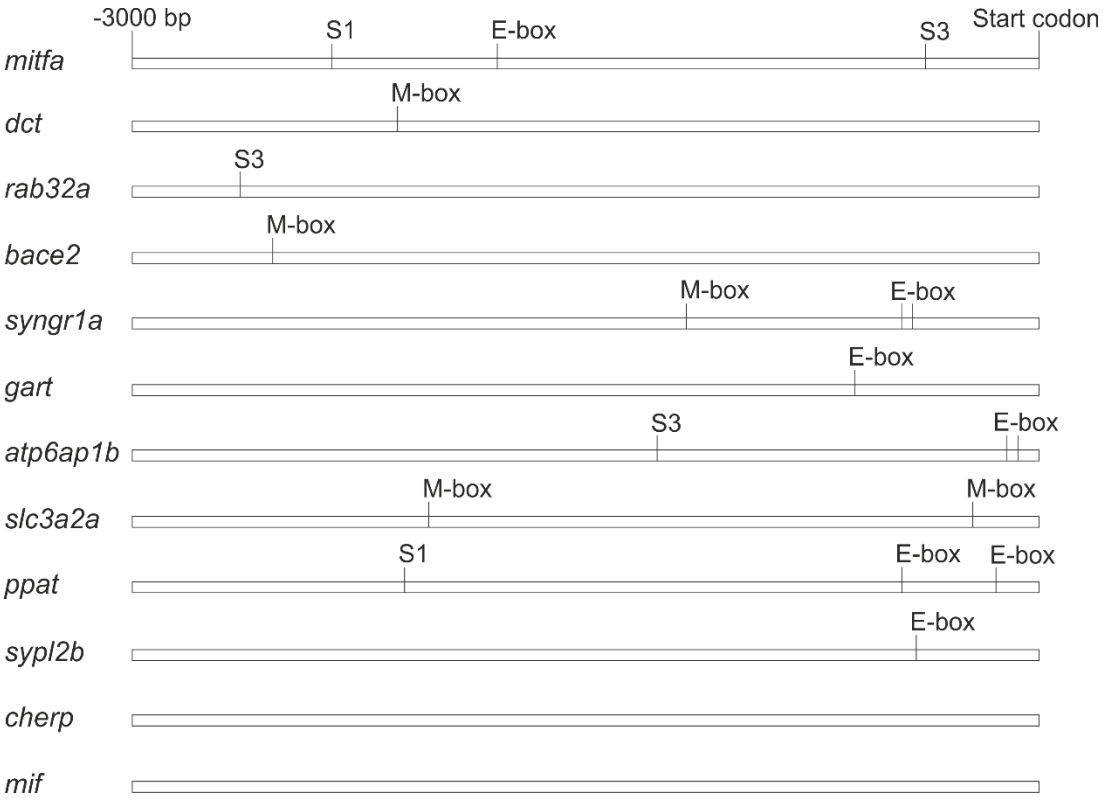

Figure 11. Representation of the promoter of the 8 genes where a *sox10* or *mitfa* binding site was found, as well as the two positive (*mitfa* and *dct*) and negative (*cherp* and *mif*) controls. The two *sox10* binding sites are labelled as S1 and S3. The two *mitfa* binding sites are labelled as M-box and E-box. The stripes represent a region of 3000 bp of the promoter upstream of the start codon, which is located on the right hand side of the stripes. The stripes and location of the binding sites within are drawn to scale.

Both the *mitfa* and *dct* promoters that were used as positive control contained Sox10 and Mitfa binding sites. The *mitfa* promoter contained two Sox10 binding sites as demonstrated by Elworthy *et al.* (2003), and the *dct* promoter contained a Mitfa binding site as established by Ludwig *et al.* (2004). Both negative control genes, that were not meant to be regulated by Sox10 or Mitfa, did not contain any Sox10 or Mitfa binding sites. This proves that the Perl scrip can successfully find transcription factor binding sides in the promotor of genes that are supposed to be regulated by Sox10 or Mitfa.

From the 13 genes only 8 contained one of the 2 binding motifs for Sox10 or Mitfa in their promoter. Three genes contained at least one of the two Sox10 binding sites. Ppat is the only one with the S1 site, and *rab32a* and *atp6ap1b* contain the S3 site, which according to Elworthy *et al.* (2003) is a weaker motif.

Seven genes contain one of the two Mitfa motifs. *bace2* has the M-box motif, but it is quite far from the translation initiation codon. *gart* and *sypl2b* also contain just one Mitfa motif (E-box), closer to the initiation codon. *syngr1a, atp6ap1b, slc3a2a* and *ppat* contain at least 2 Mitfa binding sites.

atp6ap1b and *ppat* are the only two genes containing binding sites for both Sox10 and Mitfa.

Discussion

Microarray analysis using R

Chipperfield used the freely available online tool GEPAS to analyse the microarray data he produced. This tool is based on the Bioconductor package in R, but provides a very user friendly interface, where the user can select options with buttons and drop down lists, instead of having to deal with writing an R script and using the command line in the computer. Therefore, it is a very powerful tool for biologists without a bioinformatics background. The tool works by uploading the files with your microarray results to the webpage, selecting the analysis you want to perform following the workflow, and finally you obtain the results. Unfortunately, this online tool is no longer maintained by the authors, and it is not available any more. This is why when we decided to use Chipperfield's data to try to reproduce his analysis we had to look for another option.

Coming from a bioinformatics background I have experience in analysing microarrays with the Bioconductor package for the R programming language. And, as the GEPAS tool used by Chipperfield was based in that same software, we decided to run the same analysis using the same settings, but designing our own R script.

The analysis was carried out using the 6 microarrays generated from the 7.2 transgenic line. In order to detect the differentially expressed genes between *sox10* mutant and wild type embryos the t-test statistical test was used. This test generates a list of genes ranked by t-statistic with genes downregulated in mutant embryos

having negative t-statistics values and genes upregulated in mutant embryos having positive t-statistic values. The higher the value, the more up or downregulated the gene is compared to the gene expression in wild type embryos. A p-value is produced for every t-statistic to measure the significance level of the test. The p-value measures the probability of the test statistic being equal or greater than the observed result under the assumption that the null hypothesis (no difference between wild type and mutant embryos) is true. The t-test may, however, have low power when the sample size is small, as the error variance (used to do the t-tests) is hard to estimate and subject to fluctuations when sample sizes are small (Cui and Churchill, 2003).

Two types of errors can be committed when a hypothesis test is done. The false positive error happens when a gene is declared to be differentially expressed when it is not in reality. The false negative error occurs when a true differentially expressed gene is not detected by our analysis (Cui and Churchill, 2003). Usually a statistical test to control for the false positive errors is carried out in the analysis. In microarray experiments, we perform thousands of statistical tests, one for each gene, so a large number of false positives will accumulate. Therefore, a correction for multiple testing should always be done. When Chipperfield applied the correction methods for multiple testing in his analysis no significant differentially regulated genes (with a p-value≤0.05) were detected, this includes some genes that are well known to be downregulated in *sox10* mutant embryos by *in situ* hybridisation such as *mitfa* or *dct*. This may be because the sample size is not large enough (they were comparing three WT versus three mutant arrays), as the performance of the t-test improves when increasing the sample size (Cui and Churchill, 2003). Therefore, the multiple testing correction was ignored in the analysis.

Chipperfield used an ISH screen to validate the top 100 downregulated genes he obtained from the t-test (Table 5). From those 100 candidates, he investigated 89 genes, as some were duplicated array probes for the same gene, and some ISH probes couldn't be synthetised. From the 89 genes that were assessed, 25 were validated as downregulated in pigment cells or otic vesicle in *sox10* mutants. This means that 72% of the genes identified as downregulated in the microarray analysis

were false positives (64/89). This high false positive rate is likely to be caused by the small sample size of microarrays analysed, because ISH experiment is not sensitive enough for validation purposes, as some of these genes might have been downregulated in *sox10* mutant embryos but ISH is not sensitive enough to detect it, and by not performing a false discovery correction in the analysis.

In any case, Chipperfield identified 5 novel genes as downregulated in *sox10* mutants expressed in the melanocyte lineage that did not have a published expression pattern before. These genes are regulated by *sox10* and are likely to play an important role in melanocyte biology: *atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36* (Figure 40 Appendix 2). The rest of the 20 genes were validated as expressed in the xanthophore lineage and the otic vesicle.

The list of the top 100 downregulated genes obtained through my analysis using the R script was the same as the one that Chipperfield obtained using the GEPAS online tool (Table 6). Initially both lists looked different but this was because some of the gene names in the Chipperfield's list were out of date. The zebrafish genome assembly has been updated several times since 2009, and many more genes are now annotated. Therefore, we had to search the new gene name for some of the genes in the Chipperfield's list in the ZFIN database. Once all the genes were annotated with their current name we quickly identified that both lists have the same genes. This is not surprising because we were essentially using the same Bioconductor functions that the online tool used. However, it was reassuring to know that we were capable of reproducing the initial analysis with the most up-to-date algorithms. This means we could confidently carry on using R for the rest of the analysis.

Sox10 dependent genes are also Mitfa dependent

Chipperfield identified and validated 5 novel melanocyte genes as downregulated in *sox10* mutants: *atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36*. The expression of these genes is regulated by Sox10, and therefore they are downstream of Sox10 in the melanocyte GRN. However, while *sox10* is mainly expressed in premigratory and migrating NC at early stages of development and is downregulated as melanocytes

differentiate (Kelsh, 2006b), *mitfa* is a marker of melanocyte cell fate: it drives the expression of melanogenic enzyme genes and it is also required for melanocyte survival and differentiation (Elworthy *et al.*, 2003). Therefore, we wanted to know if these genes are regulated by Mitfa, which itself is directly activated by Sox10 (Elworthy *et al.*, 2003). For that an ISH was done for each of these 5 genes in *mitfa* mutant embryos, and compared the expression pattern to the one in wild type embryos. The expression of all the genes was reduced or absent in *mitfa* mutant embryos, which means that they may have a role in melanocyte differentiation. Although we would need a mutant for each gene to test their role in differentiation, we may understand a bit more about their role in melanocyte biology by investigating their published function.

Both *atp6v1ab* and *atp6v1e1b* encode V1 subunits of the V-ATPase ion transporter. V-ATPase is the main H+ pump in animal cells, it is a multi-subunit complex where the V1 subunit catalyses the ATP hydrolysis. This protein transports H+ ions to generate a chemical gradient that contributes to vital functions in several cell processes such as the acidification of endosomes and lysosomes (Monteiro *et al.*, 2014). This protein is expressed in melanosomes, the organelles where melanin is synthesised, stored and transported. It was initially thought that an acidic environment in melanosomes facilitates melanogenesis. However, later observations suggested that some of the melanogenic enzymes are inhibited in an acid environment. It is now thought that acidification is important in the development and maturation of melanosomes. These organelles accumulate fiber-like structures in their lumen called striations where melanin is deposited. The formation of these striations requires an acidic environment in immature melanosomes. Later on, mature melanosomes lose their V-ATPase activity, favouring melanin synthesis (Tabata *et al.*, 2008).

degs1 is a dihydroceramide desaturase, the last enzyme in the de novo biosynthesis pathway of ceramides. This protein is required for the initiation of the meiosis in spermatogenesis. It was proposed that DEGS1 is a component of the contractile ring in the anaphase and telophase stages of meiosis in the spermatocyte, where its protein dysfunction leads to male sterility: DEGS1 mutants are unable to condensate

the chromosome in their nuclei. DEGS1 is also involved in apoptosis, autophagy, cell cycle regulation and lipid homeostasis (Rodriguez-Cuenca *et al.*, 2015). Its role in melanocyte development is still unclear.

pah codes for the phenylalanine hydroxylase enzyme. Tyrosinase uses tyrosine as one of its substrates to synthesise melanin, and the concentration of tyrosine depends on the conversion of the essential amino acid phenylalanine by the Pah enzyme (Schallreuter *et al.*, 2008). Thus, *pah* is a key gene in melanocyte biology.

tspan36 is a member of the tetraspanin superfamily of proteins that regulate adhesion, migration, proliferation and cell–cell fusion. Tspan36 mutation is the cause of the *dali* mutant phenotype, which shows broadened adult stripes of xanthophores and melanophores, similar to the *jaguar* phenotype. Tspan36 contributes to melanophore motility and also mediates interactions between melanophores and xanthophores in zebrafish (Inoue *et al.*, 2014).

All these genes, but *degs1* whose function is still unclear, have a clear role in melanocyte biology. They were all downregulated in *sox10* and *mitfa* mutant embryos, which means that their expression is *mitfa* dependent. Thus, these genes were placed downstream of the Mitfa transcription factor in the melanocyte GRN (Figure 12), although we do not know if their regulation is direct or indirect.

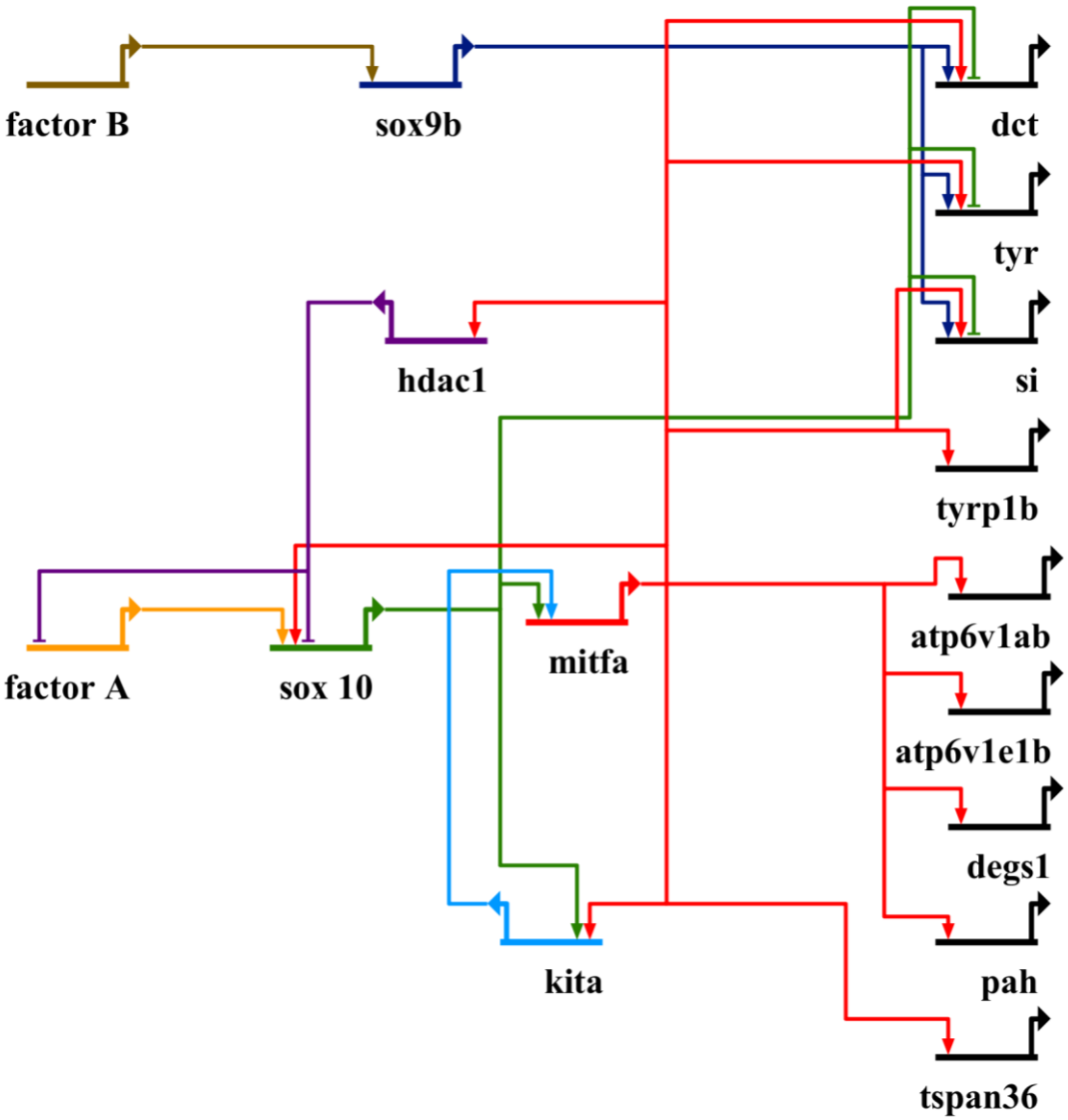

Figure 12. GRN underlying melanocyte development. This expanded network includes the core GRN published by Greenhill *et al.* (2011) and the five novel genes identified in the microarray experiment as regulated by *mitfa*: *atp6v1ab*, *atp6v1e1b*, *degs1*, *pah* and *tspan36*.

Analysis of the combined microarray dataset using improved script

In microarray analysis, as in all statistical testing, the higher the sample size the more powerful the statistical analysis is. Therefore, a higher number of microarrays will make the analysis more powerful, provide better results and reduce the number of false positives. Besides the 6 arrays that Chipperfield hybridised from the 7.2 line, he

also processed 4 more from the 4.9 line, although he did not used the later ones for the ISH screen due to time constraints. Hence, our aim was analyse the combined microarray dataset (10 arrays) to increase the power of the analysis and obtain better results.

For this analysis a modified and improved R script was used. After pre-processing the 10 microarrays together to eliminate the variability between arrays due to non-biological factors, a filtering step was carried out. This filter eliminated from the analysis the array probes that are absent in all the 10 arrays. A probe absent in all the arrays, both from wild type and mutant embryos, may indicate that the gene is not expressed in the FACS sorted cells, or that there was some issue with the array hybridisation. However, this will not be confused with a gene that is genuinely downregulated, as it will be absent in both wild type and mutant embryos.

In order to detect the differentially expressed genes, the t-test was used in the original analysis. The main issue with the t-test is that the sample variance, which is estimated using the expression values of the gene replicates, is used to calculate the t-value. With few replicates, this estimate may fluctuate considerably, making the t-value variable and unreliable. Instead of the t-test we can use linear models to assess the differentially expressed genes (Cui and Churchill, 2003).

This approach requires two matrices to be specified. The first is the design matrix which indicates which phenotypes (wild type or mutant) have been applied to each array. The second is the contrast matrix which specifies which comparisons we would like to make between the microarray samples (wild type *versus* mutant). For assessing differential expression, the empirical Bayes (eBayes) method is used (Ritchie *et al.*, 2015). eBayes finds a compromise between the variance estimated for the gene under consideration and the average variance from all the genes, giving a more reliable result with a smaller sample size and reducing the number of false positives (Ritchie *et al.*, 2015).

As said, in a microarray experiment we conduct thousands of statistical tests, one for each gene, and a substantial number of false positives may accumulate. To address this problem, there are some multiple testing correction methods that we can apply

to our analysis. The method we applied in this analysis is called False Discovery Rate (FDR), which is the proportion of false positives among all of the genes initially identified as being differentially expressed. The FDR is typically computed after a list of differentially expressed genes has been generated. It uses information available in the data to estimate the proportion of false positive results that have occurred. In a list of differentially expressed genes a proportion of these will represent false positive results at a defined FDR criterion. This method can achieve more power in reducing the false positives than other multiple testing correction methods (Cui and Churchill, 2003). As we applied an FDR correction, this analysis should contain a lower number of false positives, and therefore more true positives, than the initial one, where no multiple correction method was applied.

After filtering, applying the FDR correction and selecting the significantly downregulated genes with a p-value≤0.05 we obtain a list with 113 genes (Table 2). This list contains 45 genes from the initial list of the top 100 downregulated list obtained the initial analysis (using the 7.2 dataset and the t-test statistical method). The new list contains four out of the five genes that were validated as downregulated in *sox10* and *mitfa* mutants by ISH: *atp6v1ab, atp6v1e1b, pah* and *tspan36*. The only gene that was validated but was not found on the list is *degs1*. Degs1 is an enzyme involved in fatty acid metabolism, whose role in melanocyte biology is still unclear. So it may be that it is not involved in melanocyte development despite being downregulated in *sox10* mutant embryos and detected in our analysis of the 7.2 transgenic line arrays. Hence, this might be a case of false negative gene that was found in the analysis of the 7.2 dataset but was rejected in the new analysis incorporating the combined dataset; or it might really have a function in neural crest derived cells that is still unknown and will required further investigation. An ISH screen should be done to validate these genes.

The rest of the genes (68) have not been tested through ISH for validation yet. Thus, we cannot be sure if this analysis using the combined dataset and linear models to assess the differentially expressed genes is more powerful and produces less false positives than the initial analysis. However it seems that from the 48 genes that were also present in the 7.2 dataset analysis and were tested through ISH only four were

validated. This seems to indicate that the rate of false positives might be very similar to the initial analysis. An ISH screen would be needed to validate these genes. Nevertheless, it may be that ISH is not sensitive enough to perform this validation, so we would suggest to use quantitative PCR to validate these genes by checking if their expression is absent or downregulated in *sox10*:EGFP cells sorted by FACS from *sox10* mutant and wild type embryos.

ZFIN expression pattern of novel candidate genes

In our new analysis, 68 out of 113 genes were not investigated in the ISH screen carried out by Chipperfield in 2009. Therefore, these genes should be validated though a similar ISH screen. However, this was not possible due to time constraints. Consequently, the approach we followed was to use the ZFIN database to look for the annotated expression structures and ISH images available for each gene. Some genes are very well annotated because they have been thoroughly studied: the anatomical structures where they are expressed as well as images of their ISH expression pattern are available in ZFIN. Other genes have just the annotated anatomical structures or the ISH image. And some genes do not have any information because they have not been studied at all.

Most of the genes in our list are expressed in several anatomical structures in the embryo. We were especially interested in the genes expressed in structures with "neural crest" or "pigment cells" tags, and we labelled as "others" in Table 8 where expression was noted in any other structures. We also analysed the ISH images available and provided a description of their expression pattern. Unfortunately, some of the ISH images available in ZFIN had a low resolution, but they helped us to get a better understanding of the pattern described in the anatomical structure tags. There were 7 genes annotated as expressed in pigment cells, and their expression pattern in the ISH images looks similar to that of a pigment cell expressed gene: *rab32a, bace2, syngr1a, ednrba, gart, atp6ap1b* and *pttg1ipb.*

Rab32a is a protein that regulates endosomal trafficking. This protein is required for the formation and maintenance of notochord vacuole together with the acidification

role performed by the H+ATPase family of proteins (Ellis *et al.*, 2013). It has a role in controlling the transport of vesicles that contain the melanogenic enzymes to the melanosomes, the organelles where melanin is synthetised and stored (Raposo and Marks, 2007).

Bace2 is a type I transmembrane aspartyl proteases. In humans BACE1 is involved in the APP cleavage, which is consecutively cleaved by γ -secretase to generate Aβ, which is the main component of amyloid plaques that cause Alzheimer's Disease (according to the Aβ hypothesis). Until now, very little information has been available regarding the function of its homologue BACE2 in human. However, as therapeutic inhibition of BACE1 might also inhibit BACE2, knowledge of the physiological requirement of both protease is crucial. In *bace2* mutant embryos the migration of melanocytes is disrupted (van Bebber *et al.*, 2013). The authors suggest that the abnormal migration pattern of melanocytes resembles the *parade* zebrafish pigment mutant, although in bace2 mutants iridophore migration is not affected. The underlying molecular mechanism of the pigment phenotype is still unclear and requires further investigation.

Syngr1a is a transmembrane protein that belongs to the synaptogyrin family. In humans SYNGR1a shows mainly neuronal expression. They are involved in vesicle trafficking, but not much more is known (Kedra *et al.*, 1998).

ednrba codes for an endothelin receptor that recognises the endothelin ligand. Endothelin receptor signalling is involved in the proliferation and differentiation of neural crest-derived cells including melanocytes and their precursors, and in promoting the dendricity of fully differentiated melanocytes. It is also expressed by the iridophore and xanthophore lineages (Opdecamp *et al.*, 1998; Sviderskaya *et al.*, 1998). Zebrafish *ednrba* mutant embryos do not exhibit pigment pattern defects, pointing to a possible nonessential role in the development of pigment cells at early stages, although they have an adult pigment phenotype (*rose* mutant) In contrast in mouse, Ednrb is required by virtually all melanocyte precursors such that a failure of Ednrb signalling results in a near complete absence of melanocytes and melanoblasts (Parichy, Mellgren, *et al.*, 2000).

gart encodes an enzyme that is part of the de novo purine biosynthesis pathway together with *ppat, pfas* and *paics*. Gart encodes a trifunctional enzyme that catalyzes steps 2, 3 and 5 of IMP synthesis. *gart* and *paics* homozygous mutants are microphthalmic and they possess pigmentation defects in which nearly all xanthophore-derived yellow pigmentation and iridophore-derived silver/reflective pigmentation is absent (Ng *et al.*, 2009). Melanophore/RPE-derived black pigmentation is present but embryos appeared lighter than their wild-type siblings, possessing less melanin (Ng *et al.*, 2009). The pigmentation phenotype is due to their inability to synthesise the pigment molecule, rather than a defect on specification or migration, as transmission electron microscopy quantification of melanin levels in mutant embryos reveals a reduction in the pigment molecule. Gart and Paics function in IMP synthesis upstream of the ATP and GTP synthesis pathways. The ATP pathway modulates retinoblast proliferation and cell cycle duration within the developing vertebrate eye (Pearson *et al.*, 2002, 2005). The GTP pathway is involved in zebrafish pigmentation, probably serving as a precursor for the biosynthesis of each of the pigment molecules (Le Guyader *et al.*, 2005; Ziegler, 2003).

atp6ap1b codes for a V-ATPase-associated protein. The vacuolar ATPase (V-ATPase) is a proton pump that creates an acidic medium, necessary for lysosome function and vesicular traffic as well as for several other developmental processes. It has multiple subunits, each one performing a specific function required to achieve full activity. The zebrafish V-ATPase has 15 different subunits and mutations in any of them induce hypopigmentation or pigment dilution phenotype. V-ATPase function is essential for melanosome biogenesis, as the lack of such organelles cause the pigment dilution phenotype and may also induce melanocyte degradation, which is observed in the form of melanin spots in the zebrafish pigment dilution mutants (Ramos-Balderas *et al.*, 2013).

pttg1ipb codes for the pituitary tumour-transforming 1 interacting protein b. Not much is known about the role of pttg1ipb in melanocyte biology, however more is known about the related protein *pttg1*. PTTG1 also known as securin, participates in cell cycle regulation, DNA repair, gene transcription, apoptosis and metabolism. It plays a crucial role in the regulation of sister chromatid separation during mitosis.

PTTG1 is considered as an oncogene because of its effects on tumour development and growth. PTTG1 is abundantly expressed in multiple cancer cell lines such as melanoma (A375 and GL-Mel). High PTTG1 levels are correlated with higher tumour grade, invasiveness, and tumour vascularity (Tong and Eigler, 2009; Caporali *et al.*, 2012).

All the previous 7 genes were annotated as expressed in pigment cells in the ZFIN database, however after some investigation into their function the role in melanocyte biology of two of them (*syngr1a* and *pttg1ipb*) remains unclear. In addition to the previous genes, there are 6 other genes that are annotated as expressed in the neural crest and whose ISH expression pattern image suggests that they may be expressed in pigment cells: *pcdh10a, slc3a2a, ppat, rab7, cyb5a* and *sypl2b*. We studied these genes further to check if they may have a function in pigment cell biology.

pcdh10a codifies for the protocadherin 10a. The cadherins are transmembrane glycoproteins responsible for calcium-dependent cell-to-cell adhesion. The cadherins constitute a large superfamily that includes the classic cadherins, desmosomal cadherins, protocadherins, and atypical cadherins. Pcdh10 might serve as another cell adhesion factor in paraxial mesoderm (PAM) and somite development. *pcdh10* was also expressed in other parts of the embryo, such as epiphysis, hindbrain or otocyst epithelium, but pigment cell expression has not been reported (Murakami *et al.*, 2006).

Slc3a2a is a solute carrier family 3 member 2 (Slc3a2a). This membrane protein plays a critical role in yolk syncytial layer (YSL) formation and YSL organization by regulating microtubule networks in the YSL. Expression of *slc3a2a* has been investigated at stages of embryo development prior to 24hpf, and no pigment cell function has been reported, although they did not focus on this (Takesono *et al.*, 2012).

ppat codes for the first enzyme in the IMP de novo biosynthesis: phosphoribosyl pyrophosphate amidotransferase. *ppat* together with *gart, pfas* and *paics* are involved in the synthesis of the pigment molecules precursors (Ng *et al.*, 2009).

Therefore, it is a key enzyme in melanogenesis and its defect is likely to hinder melanocyte function.

Rab7 is a member of the Rab small glutamyl transpeptidase (GTP)-binding protein family and it is essential for the regulation of vesicular transport, organelle motility, phospholipid signalling, phagocytosis and autophagocytosis. Both tyrosinase and Tyrp1 (key enzymes in the production of melanin) are transported from the trans-Golgi network to melanosomes through endosomal compartments. Rab7 is a crucial regulator for Tyrp1 sorting in the endosomal compartment, promoting its exit from the degradative process (Hida *et al.*, 2011). Rab7 is also involved in melanoma, as the RAB7-controlled vesicular trafficking was found to be hyperactivated at very early stages of melanoma development (Alonso-Curbelo *et al.*, 2015).

Cyb5a is a membrane-bound cytochrome that reduces methemoglobin to ferrous haemoglobin, and defects in this gene cause type IV methemoglobinemia. Additionally, it functions as an electron carrier for several oxygenases and plays a key role in the detoxification of arylhydroxylamines in mammals (Giovannetti *et al.*, 2014). No link has been reported with pigment cell biology to date.

Sypl2b is member of the synaptophysin (SYP) family, which comprises integral membrane proteins involved in vesicle-trafficking events. The presynaptic SYP protein controls neurotransmitter release, while SYP-like 2 (SYPL2) contributes to maintain normal Ca2+-signalling in the skeletal muscles. While SYPL2 has a role in muscular excitation–contraction coupling in mammals, no role has been described in pigment cell biology (Andersen *et al.*, 2015).

From the previous 6 genes only *ppat* and *rab7* had a reported role in melanocyte biology, the rest perform important but apparently unrelated roles to pigmentation.

Location of Sox10 or Mitfa binding sites in candidate genes

All of the previous 13 genes were found to be downregulated in *sox10* mutant embryos in the microarray analysis, therefore they are regulated (directly or not) by

sox10 transcription factor. Some of these genes may be part of the melanocyte GRN, but they might also have a role in other pigment cells biology (iridophore or xanthophore) or other *sox10* expressing cells (glia, oligodendrocytes or sensory neurons). The analysis of their function from the published literature suggests that seven of them have a function in melanocyte biology. Hence, it will be helpful to know it they are directly regulated by *mitfa* (the melanocyte master regulator), as this would indicate that they are downstream of *mitfa* in the melanocyte GRN.

To find out if these 13 genes are regulated by Mitfa we downloaded from the Ensemble zebrafish genome repository a region of 3000 bps of the promoter and 5'UTR of the 13 genes. Subsequently, we built a Perl program to find the Mitfa (and Sox10) transcription factor binding motifs on these gene's promoters. There are programs available to find TF binding sites (ConTra or JASPAR), but all of them use binding motifs from ChIP experiments, so only those TF that have been investigated through this technique will have a binding motif in the database, and the program will be able to find them in the candidate promoters. Because Mitfa does not have a transcription factor binding motif for zebrafish in the database, we could not use the available programs, so we decided to build our Perl program.

mitfa is known to be directly regulated by Sox10 (Elworthy *et al.*, 2003), and *dct* has been identified as regulated by Mitfa (Ludwig *et al.*, 2004). Hence, we used the promoters of both of these genes as a positive control to test if our Perl program could find transcription factor binding sites. As expected the program found that the *mitfa* promoter contained Sox10 binding sites, and the *dct* promoter contained Mitfa binding sites. However, we still needed to check that our program was finding true positives and it was not finding binding sites for Sox10 and Mitfa in any random gene. For that we selected two genes from the list of differentially expressed genes obtained from the microarray analysis whose expression was not statistically different between *sox10* mutant and wild type embryos (both genes presented a p-value close to 1). The genes selected were *cherp* and *mif*. *cherp* is a calcium homeostasis endoplasmic reticulum protein (Gene ID: 10523) and *mif* is a macrophage migration inhibitory factor (Gene ID: 4282). The function of both of these genes is not related to melanocyte biology, and hence it was not expected to

find Sox10 or Mitfa binding sites. As predicted our Perl program did not identify any Sox10 or Mitfa binding sites in the *cherp* and *mif* promoters, which confirmed that our program was successful in finding real binding sites.

From the 13 genes analysed 8 contained Sox10 or Mitfa binding sites in their promoter. *Rab32a, atp6ap1b* and *ppat* are the only ones containing Sox10 binding sites. Seven genes contain one of the two Mitfa motifs. Bace2 has just one M-box motif, and Gart and sypl2b also contain just one Mitfa motif (E-box). *syngr1a, atp6ap1b, slc3a2a* and *ppat* contain at least two Mitfa binding sites. *Atp6ap1b* and *ppat* are the only two genes containing binding sites for both Mitfa and Sox10.

Rab32a has an important role in melanocyte biology, as it controls the transport of the melanogenic enzymes to the melanosomes. However its promoter contains just a Sox10 S3 binding site, which is weaker driving expression than S1 (Elworthy *et al* 2003).

Bace2 may be involved in melanocyte migration as it has a Mitfa M-box site in its promoter. The role of *syngr1a* in pigment cell biology is not well understood, however it contains an M-box and two E-boxes in tis promoter, which makes it quite likely to be regulated by *mitfa*. Nevertheless further study is needed.

Gart is an enzyme involved in the biosynthesis of the pigment molecules together with Ppat. *ppat* has a Sox10 S1 binding motif, which is the strong one (Elworthy *et al* 2003), and two Mitfa E-boxes, suggesting that it may likely to be controlled by Mitfa (or Sox10). *gart* has a single Mitfa E-box in its promoter, and as its role is also very important for melanocyte biology it may be regulated by Mitfa.

Atp6ap1b is involved in the acidification of the melanosome, which is essential for melanocyte development. It contains both Sox10 S3 and two Mitfa E-box sites. That makes it likely to be regulated by Mitfa.

slc3a2a contains two Mitfa M-boxes in its promoter, however, its role in pigment cell biology is not clearly understood. *sypl2b* contains one Mitfa E-box in its promoter. The unknown role in pigment cell biology of these genes make them unlikely to be regulated by Mitfa, although further investigation would be needed to test this.

Hence, to sum up, it is very likely that *gart, ppat* and *atp6ap1b* are regulated by Sox10 or Mitfa. It may be possible that *rab32a* and *bace2* are also regulated by Sox10 or Mitfa. In any case, experimental validation will be needed. One way to test if these genes are regulated by Sox10 or Mitfa would be to perform ISH using *sox10* or *mitfa* mutant embryos for these genes. If their expression is reduced or absent in the mutant embryos that means that are regulated by those transcription factors. However with this method we would not know if the interaction is direct or indirect. In order to test if the interaction is direct, ChIP-seq could be used. The main issue with this approach is that we need suitable anti-Sox10 and anti-Mitfa antibodies, which are difficult to obtain. The advantage is that once we have an antibody that works with our sample we perform the ChIP-seq experiment and identify all the genes that are directly bound by Sox10 or Mitfa.

Conclusion

The aim of this chapter was to use the microarray data produced by Chipperfield in 2009 to find novel genes involved in the GRN underlying melanocyte development. We repeated the analysis he did several years ago with up to date algorithms using the Bioconductor package for R, and found out that the list of top 100 downregulated genes in *sox10* mutant embryos that we obtained is the same.

Chipperfield performed an ISH screen of the those top 100 downregulated genes to validate that *atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36* are expressed in melanocytes and downregulated in *sox10* mutants. We carried out another ISH experiment to check the expression of these genes in *mitfa* mutant embryos, and found out that they are also downregulated. Therefore, those 5 genes are regulated (directly or indirectly) by Mitfa and we placed them downstream of Mitfa in the melanocyte GRN.

We used the combined microarray dataset (obtained from both Tg(-7.2*sox10*:EGFP) and Tg(-4.9*sox10*:EGFP) lines) and an improved statistical analysis to obtain a new list of downregulated genes. This list contained 68 new candidate genes that were not present in the initial analysis of the Tg(-7.2*sox10*:EGFP) line. We used the ZFIN

database to investigate the expression profile of these 68 genes, and found that 7 of them are very likely to perform a key role in melanocyte biology*: rab32a, rab7, bace2, ednrba, gart, ppat* and *atp6ap1b*; and 6 other genes that may have a role in melanocyte development, but whose function is still unclear.

Finally, we designed a Perl program to locate Sox10 or Mitfa binding sites in the promoter of these 13 genes, which may indicate that are directly regulated by one of the transcription factors. Three of the genes (*gart*, *ppat*, and *atp6ap1b*) are likely to be regulated by Mitfa or Sox10, as they contained several transcription factor binding motifs in their promoter. Two other genes (*rab32a* and *bace2*) may also be regulated by the transcription factors. In any case, this would need to be tested experimentally.

To sum up, the microarray analysis has provided five novel genes regulated by Mitfa that are important in melanocyte biology (*atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36*) and have been validated experimentally; as well as seven candidate genes that are directly or indirectly regulated by Mitfa or Sox10 (*rab32a, rab7, bace2, ednrba, gart, ppat* and *atp6ap1b*) that may have a role in melanocyte biology, which would need to be experimentally tested.

Chapter 4:

Including *kit* in the GRN

Introduction

As described in the introduction, the GRN published by Greenhill *et al.* (2011) identified the need for some unknown factors (genes) in order to describe melanocyte development adequately. One of those factors was Factor Y, which is a Sox10-independent, Mitfa-dependent transcriptional activator of *mitfa* that provides a positive feedback loop allowing stable melanocyte differentiation. Mc1R was one of the candidates for Factor Y but Greenhill *et al.* (2011) concluded that Mc1R signalling was not contributing to Factor Y.

Hence, a different gene had to be considered to act as Factor Y in the current GRN, and based on studies carried out by many research groups *kita* (termed *kit* in the rest of the chapter) appeared as a very interesting candidate to investigate. This section will describe the evidence that supports the selection of *kit* as candidate for Factor Y.

Kit is a type III receptor tyrosine kinase expressed by melanocytes and melanoblasts (melanocyte precursors) that that binds to the Stem Cell Factor (SCF, also called Steel factor) which is its ligand. Upon binding with its ligand, Kit transduces the signal through the PI(3) kinase and MAP kinase signalling cascades. MITF protein contains a phosphorylation site that is phosphorylated by ERK2, which lies in the MAP kinase pathway downstream of *kit* and other RTKs (Figure 13) (Goding 2000; Lister *et al.* 1999)

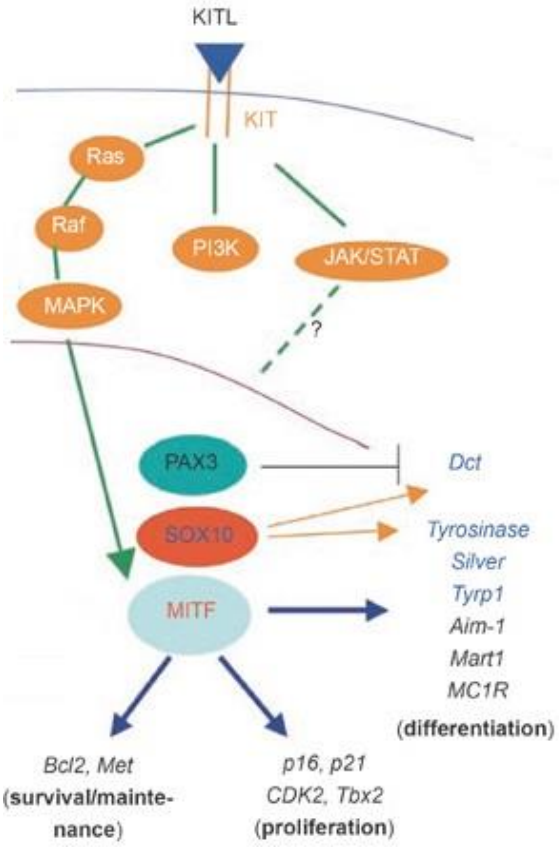

Figure 13. A simplified schematic showing the features of the *kit* signalling pathways in melanocyte development. Kit regulates Mitf activity through the MAP kinase cascade upon binding with its ligand. Modified from Hou & Pavan (2008).

Kit signalling was proposed to initiate melanoblast migration, attracting them onto the dorsolateral pathway lined by cells expressing SCF. In zebrafish *kit* mutants, melanocytes are specified and differentiate, however they fail to migrate properly and tend to accumulate in premigratory positions (dorsal to the neural tube, and in a cluster posterior to the inner ear), and ultimately they enter the epidermis where they start to synthesise melanin (Parichy *et al.* 1999; Kelsh *et al.* 2000). Using a conditional mutant, the sensitive phase for Kit activity in melanocyte migration is restricted to between 1 and 2 dpf. After that, the migration machinery of the cells and the extracellular matrix have changed, and the melanocytes cannot migrate more (Rawls and Johnson, 2003). Elegant studies of the phases of Kit requirement during murine melanocyte development used blocking antibody injections to demonstrate the early phase requirement, but also a later one for population of hair follicles (Nishikawa *et al.*, 1991). This second role of Kit is proposed to be a

chemokinetic effect: Kit increases the melanoblasts' rate of movement and, therefore, their chance of encountering the follicle (Jordan and Jackson, 2000).

A temperature shifting experiment, carried out using a *kit* temperature sensitive mutant (which acts as a null mutant when reared at the restrictive temperature of 33^{o}C) showed that *kit* is required for survival in a transient period of time between 2 and 4 dpf, as *kit*^{ts} mutant embryos reared at the permissive temperature of 25^{o}C and shifted to the restricted temperature between 2 and 4 dpf caused melanocyte death (Rawls and Johnson, 2003). Melanoblast apoptosis in Kit mutants results from a failure of the receptor tyrosine kinase to activate the PI(3) kinase and MAP kinase cascades. PI(3) kinase delivers a powerful anti-apoptotic signal via the activation of the kinases PDK1 and Akt (Goding, 2000). Another target of Kit signalling is Bcl2, a known pro-survival gene (McGill *et al.*, 2002).

kit is not required for melanocyte differentiation, but it may promote that a maximum number of melanocyte are differentiated (Mellgren and Johnson, 2004; Parichy *et al.*, 1999). *kit* is also involved in the maintenance of melanocytes: in the absence of *kit* the melanocytes undergo cell death, and are extruded from the skin (Parichy *et al.*, 1999; Kelsh *et al.*, 2000).

In mouse Kit mutants, there are fewer melanoblasts expressing Dct, suggesting that Kit could be required for expansion or survival of melanoblasts. Zebrafish *kit* mutants have a lower number of melanocytes than wild type embryos, suggesting that some melanoblasts may require *kit* to develop normally (Kelsh *et al.*, 2000). In zebrafish, an experiment where *mitfa*-MO was injected into wild type embryos showed no *kit* expressing melanoblasts at 24 hpf, suggesting that *kit* requires *mitfa* for its expression (Mellgren and Johnson, 2004).

All the previous evidence suggests that *kit* is a key gene in melanocyte development, involved in melanocyte migration and maintenance, as well as promoting its differentiation in a maximum number. However the following experiments were the ones that made us hypothesise that *kit* may act as Factor Y in the Greenhill *et al.* (2011) GRN.

In a study using Mitf mutant mast cells that lacked Kit expression, transfection with a Mitf plasmid resulted in Kit expression. This was shown to be because the Kit promoter contains an E-box motif that is recognised by MITF to activate Kit transcription (Tsujimura *et al.*, 1996). In melanoblasts, MITF was also shown to be required to drive Kit transcription (Opdecamp *et al.*, 1997). This suggests the possibility of a positive feedback loop between Kit signalling and Mitf.

A different study showed that KIT-negative cells in culture express MITF and several other pigment cell genes directly regulated by MITF, but lack Tyr expression. This indicates that KIT is not required for Mitf expression nor for the expression of other pigment cell genes. However, without Kit signalling, MITF is not sufficient do drive Tyr expression in developing melanoblasts (Hou *et al.*, 2000).

Based on the previous evidence our hypothesis (Figure 14) was that *kit* is performing the same role as Factor Y (at least partially) in the GRN: its expression is driven by *mitfa*, and *kit* itself drives and maintains *mitfa* expression through a positive feedback loop.

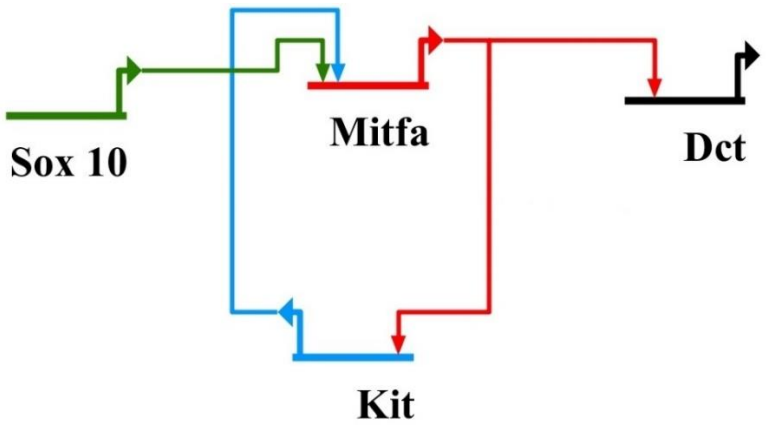

Figure 14. Hypothesis of the role of *kit* in the GRN. Our initial hypothesis stated that *kit* plays the role of Factor Y in the Greenhill *et al.* (2011) GRN as a central actor in a feedback loop with *mitfa*. Mitfa, whose expression is directly regulated by Sox10, drives *kit* expression, and *kit* itself promotes *mitfa* transcription helping to maintain its expression level to promote stable melanocyte differentiation.

The approach selected to test our hypothesis was to perform whole mount *in situ* hybridisation to detect *kit* expression in several zebrafish genetic backgrounds. In order to test if *kit* is driven by Mitfa we performed *kit* ISH in *mitfa* mutant embryos. Lack of melanocytes expressing *kit* would suggest that Mitfa regulates *kit* expression. To test whether Kit maintains *mitfa* expression we performed *mitfa* ISH in *kit* mutants. The reduction of *mitfa* expression in *kit* mutants may suggest that Kit signalling maintains *mitfa* expression.

Results

kit is driven by Mitfa

It has been shown that the Kit promoter contains an E-box motif that is recognised by MITF to activate Kit transcription (Tsujimura *et al.*, 1996). And the same time MITF was also shown to be required to drive Kit transcription (Opdecamp *et al.*, 1997). Therefore we wanted to test our hypothesis: if *kit* is factor Y in the network, it is in the centre of a feedback loop, where its expression is driven by Mitfa, and Kit *activity* itself drives or, at least, maintains *mitfa* expression. First of all I checked if *kit* expression is driven by Mitfa. For that, an ISH of *kit* was done in wild type, *sox10* and *mitfa* mutant embryos at 30hpf. If our hypothesis is right, *kit* expression should be absent in both *sox10* (that directly drives *mitfa* (Elworthy, 2003)) and *mitfa* mutants (Figure 15).

The results of the *kit* ISH in wild type, *sox10* and *mitfa* mutant embryos can be seen in Figure 15. Panel A shows that as expected, *kit* is expressed in the melanocyte lineage as well as other tissues like the proctodeum or notochord (Ignatius *et al.*, 2008). However, it was rather surprising to observe that it was also widely expressed in mucus secreting cells in the epidermis, which had not been previously identified when studying the *kit* expression pattern through ISH. These round and small cells are located in the fish epidermis in high number. They had been previously described in other teleost fish as the cells that secrete the mucus that acts as a lubricant reducing drag while swimming, preventing fungal and bacterial infection and being

involved in osmoregulation (Marshall, 1976). These cells seem to have a non-neural crest origin and hence they would not be regulated by *sox10* or *mitfa*. Thus, we expected to find them in the *kit* ISH for both *mitfa* and *sox10* mutants.

As we expected we could not see *kit* expressing melanocytes in *sox10* mutants as *mitfa* would not be active and it could not drive *kit* expression (Figure 15 B). However, *kit* was still expressed in the mucus secreting cells and in the proctodeum, as predicted, since these tissues have no neural crest origin.

In *mitfa* mutants we expected to find a lack of *kit* expressing melanocytes, as *mitfa* would not be able to drive *kit* expression, but we would still see mucus secreting cells and proctodeum expressing *kit*. Figure 15 C shows the ISH results where we could appreciate that *kit* was expressed in mucus secreting cells and proctodeum as predicted; however some there were some large cells located dorsally expressing *kit*. These cells were too big to be mucus secreting cells; therefore we hypothesised that they might be melanocytes, because they are in a place where melanocytes are usually found.

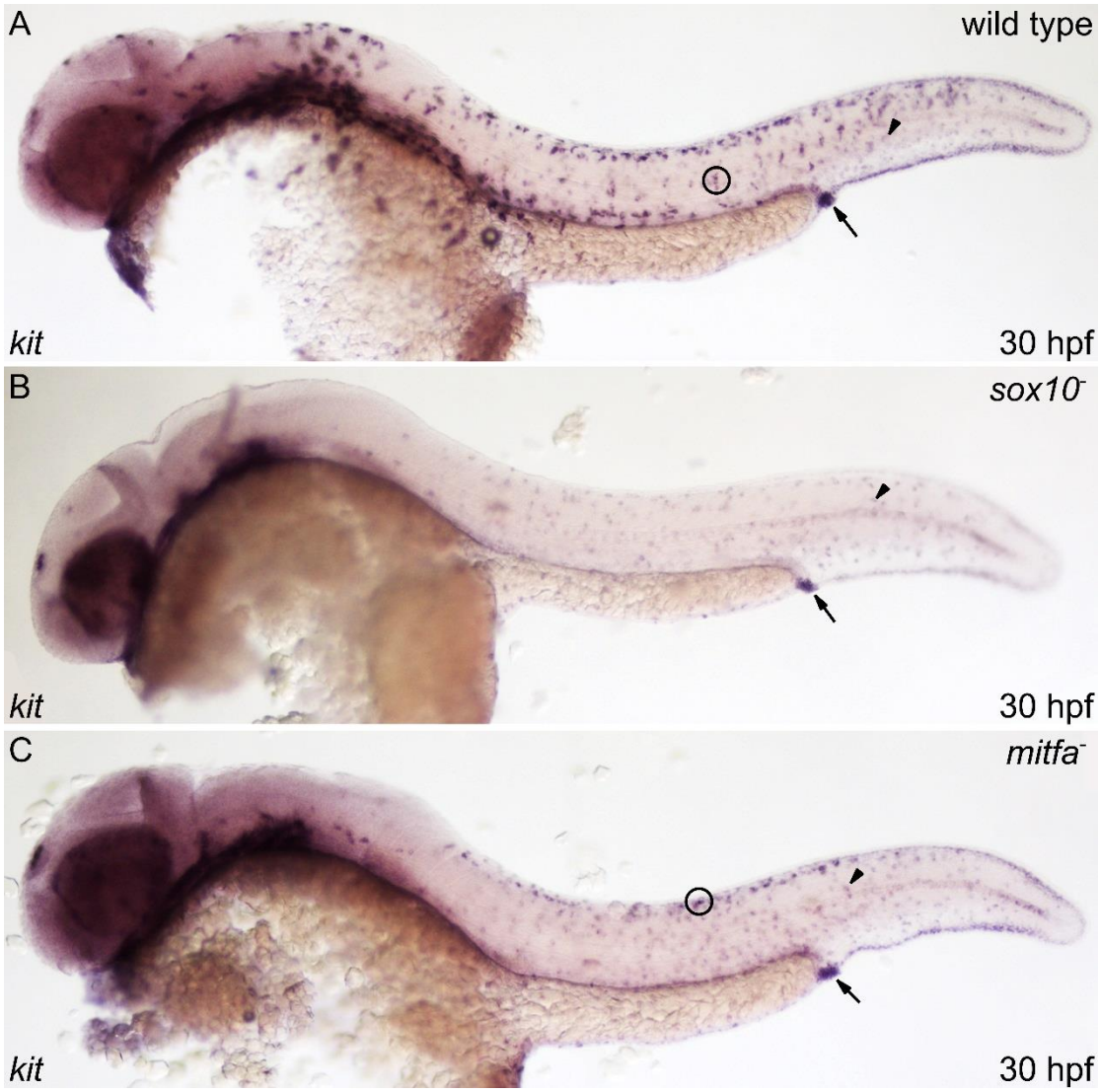

Figure 15. *kit* ISH of 30 hpf embryos shows that *kit* may be unexpectedly expressed in *mitfa* mutant embryos. Wild type embryo (A) show *kit* expression in melanocytes (circle), mucus secreting cells (arrowhead), proctodeum (arrow) and notochord (tail region), as expected. (B) The *sox10* mutant embryo shows *kit* expression in mucus secreting cells (small, round cells label with arrowhead), proctodeum (arrow) and notochord (tail region), but not in melanocytes as predicted. (C) The *mitfa* mutant embryo shows *kit* expression as expected in mucus secreting cells (arrowhead), proctodeum (arrow) and notochord (tail region), and some unexpected *kit* expressing cells dorsally (circle) that might be melanocytes. Image orientation: embryo head is on the left and tail on the right of the images.

In the *kit* ISH of *mitfa* mutant embryos it was not expected to find *kit* expressing melanocytes according with our hypothesis. However we found some cells expressing *kit* that we thought might be melanocytes. We were not sure that those cells that we believed are melanocytes were not actually mucus secreting cells. Therefore we needed to find a way to distinguish between both cell types. The main

characteristic of the mucus secreting cells is that they are located in the epidermis in the surface of the embryo, whereas the melanocytes are located a bit deeper into the embryo body in the medial plane. So we could use this feature to distinguish them under a microscope as each cell type will be in a different focal plane.

We performed a *kit* ISH for 24 hpf wild type, *sox10* and *mitfa* mutant embryos and we imaged the anterior part of the embryo with two focal planes one of the lateral side of the embryo and one of the medial pathway. Figure 16 shows that in wild type fish there are mucus secreting cells expressing *kit* in the surface (round and small cells (A), and melanoblast deeper inside the embryo (D). Sox10 mutant embryos have mucus secreting in the epidermis (B), but no melanoblasts are present deeper inside the embryo (E). However, in *mitfa* mutant embryos we could find mucus secreting cells expressing *kit* in the surface of the embryo (like in wild type and *sox10* mutants, C) and the cells that we suggested could be melanoblasts deeper inside the embryo (F), which are not present in *sox10* mutants. However this was not conclusive enough, as these cells that we thought were melanoblasts could be mucus secreting cells located in the epidermis of the dorsal side of the embryo. Hence we needed to find a different way to test this.

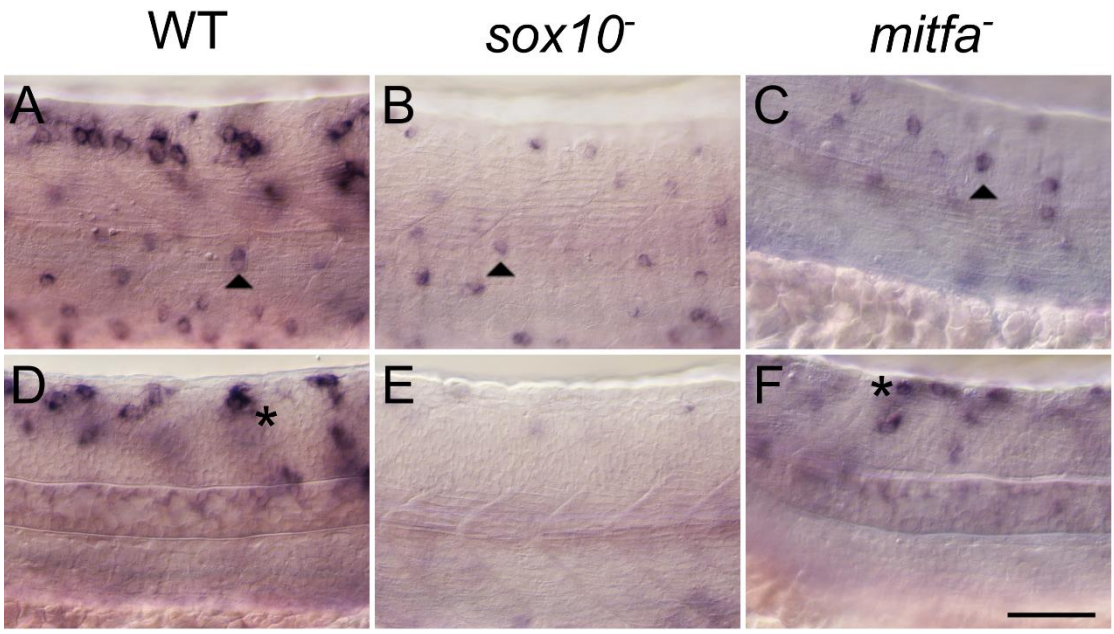

Figure 16. *kit* ISH of 24 hpf embryos shows presumptive melanoblasts in the *mitfa* mutants medial pathway.. A, B and C are images of the lateral focal plane, and D, E and F are images of the medial migration pathway of the same embryos. Wild type embryo shows *kit* expression in mucus secreting cells in the lateral plane (arrowheads) (A) and in melanoblast in the medial plane (stars) (D). *sox10* mutant shows *kit* expression in mucus secreting cells (B) but not in melanoblasts in the medial plane (E). *mitfa* mutant shows *kit* expressing mucus secreting cells in the lateral plane (C) as well as larger *kit* expressing cells in the medial plane (probable melanoblasts) (F). Scale bar: 100 µm. Images taken from the mid trunk, above the yolk sac extension. Image orientation: embryo head is on the left and tail on the right of the images.

To test the proposal that these larger *kit* positive cells in the *mitfa* mutants were melanoblasts, I used double fluorescent *in situ* hybridisation for *mitfa* and *kit*. If those *kit* expressing cells in the medial plane are melanocytes they should also express *mitfa*, and therefore we should detect co-localisation. If on the contrary they are mucus secreting cells we would see *kit* but not *mitfa* expression.

As expected *mitfa* and *kit* co-localise in melanocytes on the medial plane of wild type embryos (Figure 17 A-C). In *mitfa* mutant some cells in the typical position of premigratory neural crest cells located above the dorsal neural tube co-express both *mitfa* and *kit* (Figure 17 D-F). This suggests that the presumptive melanocytes that expressed *kit* initially detected in *mitfa* mutants are indeed melanocytes.

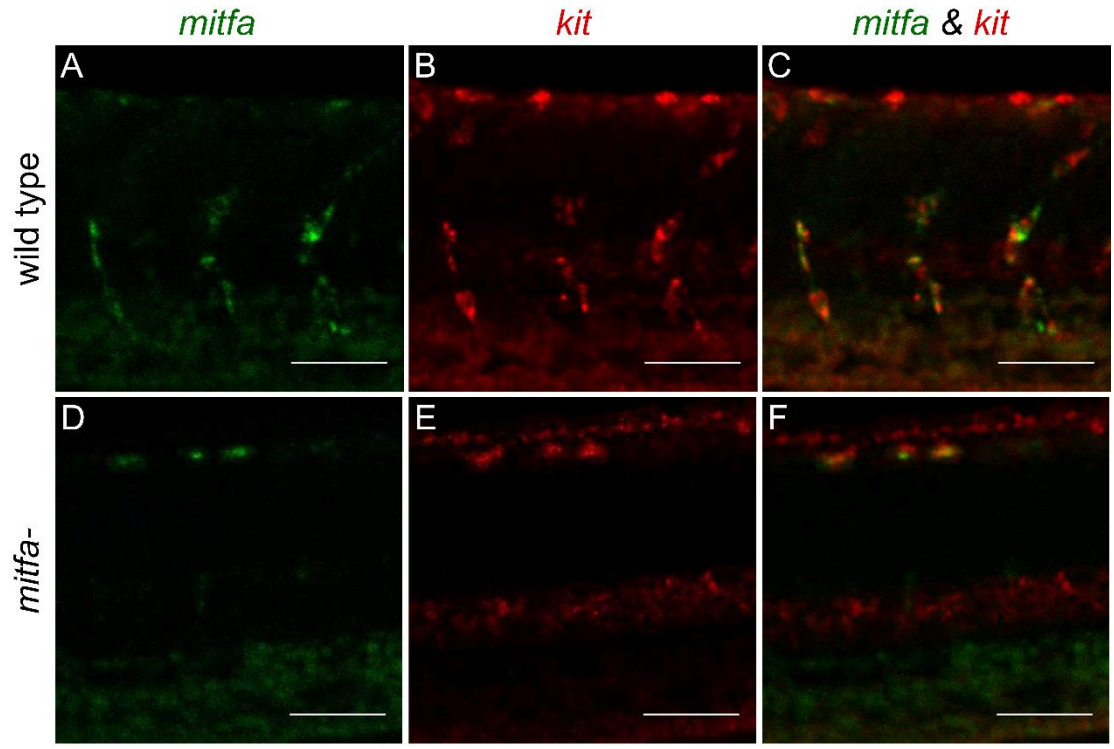

Figure 17.medial pathway images of ISH of 30 hpf wild type, *sox10* and *mitfa* mutant embryos for *mitfa* (green) and *kit* (red) suggest that *mitfa* mutants have mealnocytes in the dorsal trunk. As expected, *mitfa* and *kit* expression overlaps extensively in cells of typical neural crest morphology on both medial (A-C) and lateral (not shown) pathways. In *mitfa* mutants migratory cells are not seen, but some cells in the typical position of premigratory neural crest cells are seen above the dorsal neural tube; these cells show co-localisation of *mitfa* and *kit* (D-F). Scale bar: 100 µm. Images taken from the mid trunk, above the yolk sac extension. Image orientation: embryo head is on the left and tail on the right of the images.

Our initial hypothesis proposed that *kit* expression was driven by *mitfa*, and hence no *kit* expressing melanocytes were supposed to be found in *mitfa* mutants. However the previous data suggests that there is some other factor driving *kit* expression in melanocytes even when *mitfa* is not functional. And we proposed that this factor might be *sox10*, as *kit* expressing melanocytes are not present in *sox10* mutant. This finding led us to modify our initial hypothesis: *kit* might be driven by *sox10*, as well as being activated by Mitfa at the same time that it maintains *mitfa* expression as part of the feedback loop.

Greenhill *et al.* (2011) established that all melanocytes showed detectable *sox10* expression at 30 hpf, but this rapidly decreased and by 50 hpf signal was not detected

in any cell. Hence, if *sox10* is driving *kit* expression this must be limited to stages prior to 50 hpf, that is why we decided to evaluate the timeframe of *sox10* driving *kit* expression.

kit is transiently driven by Sox10

In order to evaluate the timeframe of *sox10* driving *kit* we performed an ISH for *kit* using three different stages (24, 30 and 48 hpf) and wild type, *sox10* and *mitfa* mutant embryos. As we had described before we assumed we would not find melanocytes expressing *kit* in *sox10* mutant embryos in any of the three stages. However, we expect to find melanoblasts expressing *kit* in *mitfa* mutants at 24 hpf, and 30 hpf. We do not expect to find melanocytes expressing *kit* at 48 hpf, as by this stage *sox10* expression in melanocytes is almost absent (Greenhill *et al.*, 2011).

We could identify melanocytes and their precursors expressing *kit* in wild type embryos in all three stages (24, 30 and 48 hpf). In addition, we could not see any melanocyte expressing *kit* in *sox10* mutant in any of the stages as expected. There were some *kit* expressing melanoblasts in *mitfa* mutant at 24 hpf, which were absent from 48 hpf *mitfa* mutant embryos as predicted. However, contrary to our predictions, we could not detect any *kit* expressing melanocyte from 30 hpf *mitfa* mutant embryos. This may be because the levels of *sox10* at 30 hpf are already decreasing and hence fewer *kit* expressing melanocytes are found at this stage.

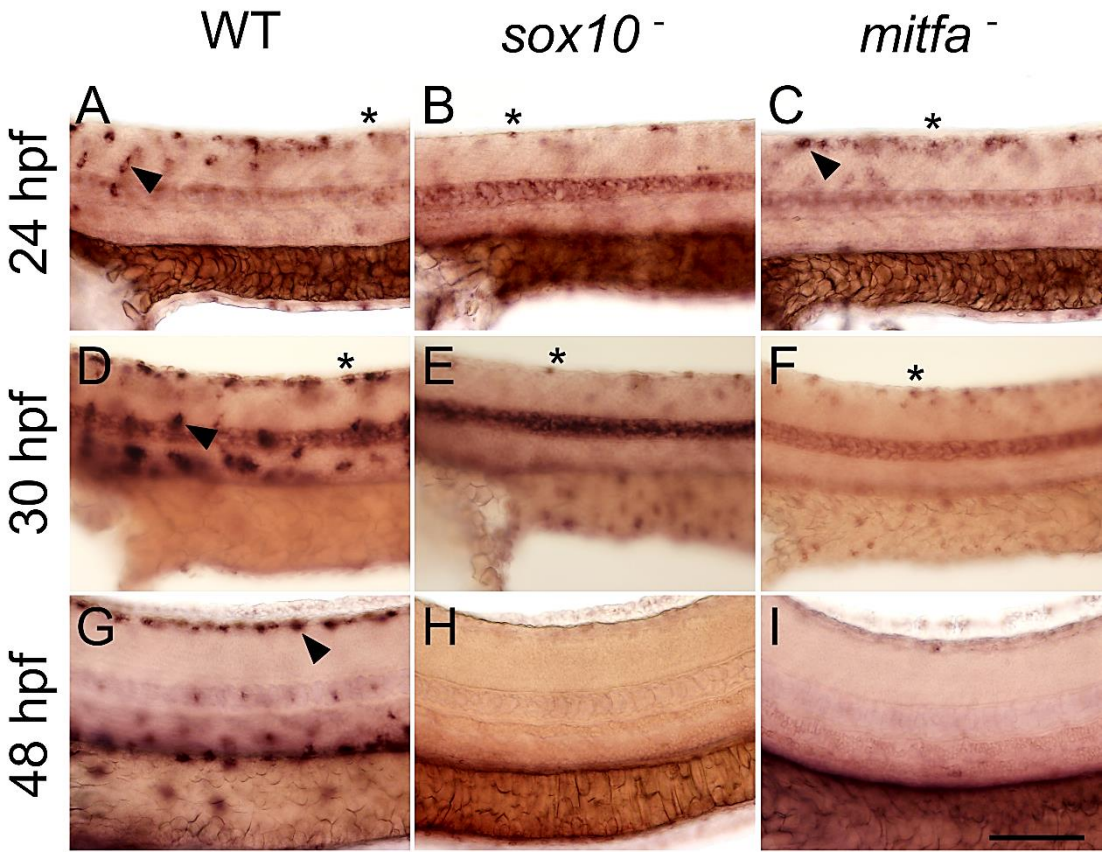

Figure 18. *sox10* may drive *kit* expression at early stages of development as *kit* expressing melanocytes were found in *mitfa* mutant embryos at 24 hpf but not at 30 or 48 hpf. At 24 hpf *kit* is expressed in mucus secreting cells (stars) and melanoblast (arrowheads) in wild type embryos (A), just in mucus secreting cells in *sox10* mutant embryos (B), and in mucus secreting cells and melanoblast in *mitfa* mutants (C). At 30 hpf *kit* is expressed in mucus secreting cells and melanocytes in wild type embryos (D), and only in mucus secreting cells in *sox10* (E) *mitfa* mutants embryos (F). At 48 hpf *kit* is expressed in melanocytes in wild type embryos (G), and it is not expressed in any cells in *sox10* or *mitfa* mutants (H-I). These ISH images are from the medial pathway focal plane and from the mid trunk, above the yolk sac extension. Image orientation: embryo head is on the left and tail on the right of the images. Scale bar: 100 µm

To sum up, the previous results demonstrate that Mitfa controls *kit* expression and *sox10* drives *kit* expression in early stages of melanocyte development (between 24 and 30 hpf). This confirms part of the interactions that are consistent with *kit* acting as Factor Y. The other part of the interaction that we need to test is *kit* driving *mitfa* transcription, maintaining its expression in a feedback loop.

Kit drives mitfa and maintains its expression through a feedback loop.

If *kit* acts as Factor Y in the Greenhill *et al.* (2011) GRN we needed to assess if *kit* drives and maintains *mitfa* expression through the feedback loop, as we have already demonstrated the other part of the interaction. To test this, an *in situ* hybridisation using *kit* mutant embryos was carried out for two genes: *mitfa* and *dct*. If *kit* plays a role in this loop *mitfa* expression levels should be decreased, as should the levels of the *mitfa* downstream target *dct*.

According to our model, both *mitfa* and *dct* expression should be reduced in *kit* mutant embryos as *kit* would not be functional to drive and maintain the expression level of *mitfa*, which would decrease the expression of its downstream targets such as *dct*. However, Figure 19 shows a reduction in the number of *mitfa* and *dct* expressing cells in *kit* mutants compared to their wild type siblings, rather than a reduction in their expression levels. From Figure 19 it seems that *mitfa* expression level might be a bit weaker in *kit* mutant than in wild type, however *dct* looks similar. Therefore, we would need a quantitative measure of expression levels, which could be obtained by RT-qPCR, to characterise this aspect of the *kit* interaction with *mitfa*.

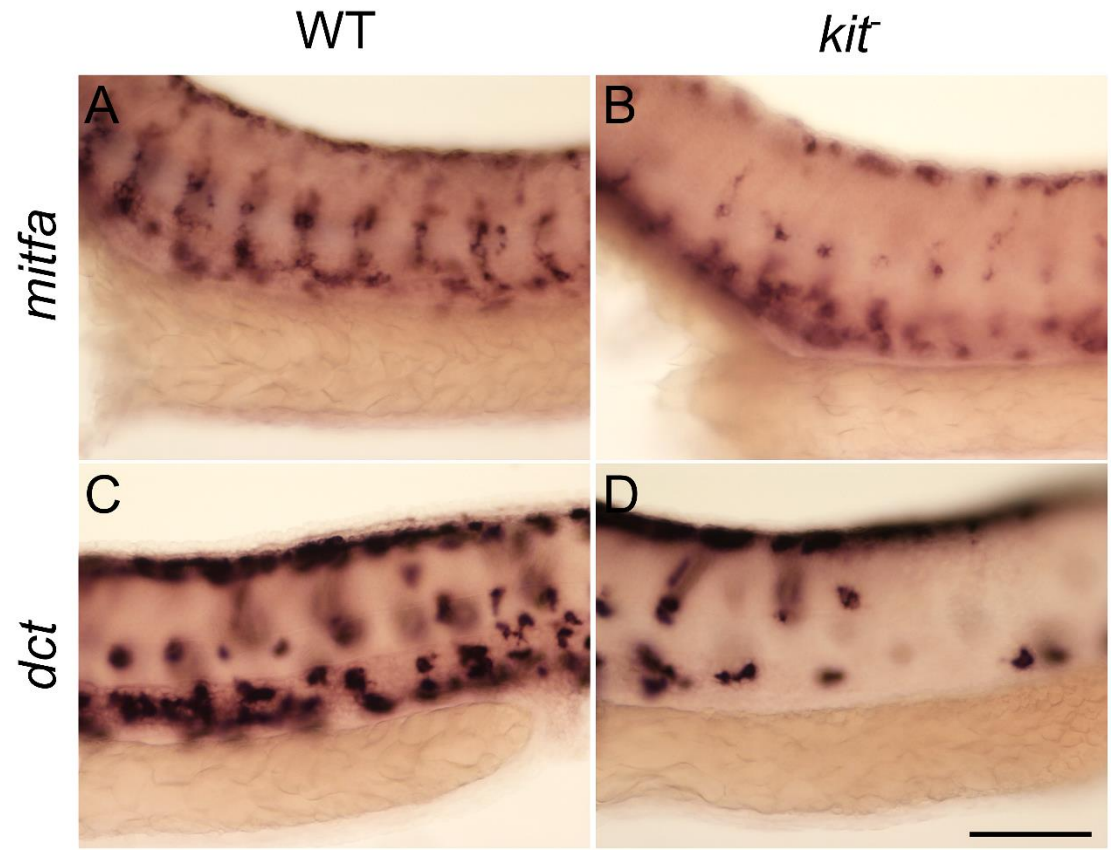

Figure 19. *mitfa* and *dct* ISH of 30 hpf wild type and *kit* mutant embryos. Kit mutant embryos show expression of both *mitfa* (B) and *dct* (D) in a reduced number of melanocytes compared to the wild type embryos (A and C). Not enough information about gene expression levels can be drawn from these ISH images. Images taken from the mid trunk. Image orientation: embryo head is on the left and tail on the right of the images. Scale bar: 100 µm

Discussion

kit is driven by Mitfa

Based on the evidence gathered from the literature review presented in the introduction our initial hypothesis was that *kit* might be playing the role of Factor Y in the Greenhill *et al.* (2011) GRN at least in part (i.e. there might be another gene or genes sharing the role of Factor Y with *kit*). According to this hypothesis *kit* would be part of a feedback loop with *mitfa*, where *kit* expression is driven by Mitfa, and Kit drives and maintains Mitfa expression. This hypothesis is represented in Figure 14.

To test this hypothesis an ISH for *kit* was done using wild type, *sox10* and *mitfa* mutant embryos. If the hypothesis is right, we would not expect to find *kit* expressing melanocytes in *sox10* and *mitfa* mutants, because *kit* transcription would not be activated in those fish. Figure 15 A shows the expected *kit* expression pattern in wild type fish. We can detect *kit* expressed in melanocytes as well as in mucus secreting cells, which had not been reported in the previous studies in zebrafish (Parichy *et al.*, 1999; Cooper *et al.*, 2009). These cells are completely unrelated to the melanocyte lineage (Marshall, 1976) and hence are not important in the process we are trying to understand. Figure 15 B shows the expected expression pattern: *kit* expressed in mucus secreting cells but not in melanocytes. However, the expression pattern of *mitfa* mutant embryos (Figure 15 C) was unexpected, as there seems to be some *kit* expressing cells that are bigger than mucus secreting cells, which may be melanocytes. In order to characterise this properly a more detail observation was done. Mucus secreting cells are located in the epidermis in the external layers of the embryo, whereas melanocytes are located deeper within the embryo (Kelsh 2004). Therefore we could use the two different focal planes in the microscope to elucidate if those cells that might be melanocytes in *mitfa* mutant embryos are indeed that.

Figure 16 shows an ISH for *kit* using wild type, *sox10* and *mitfa* mutant embryos with images of the lateral and medial planes of each of them. In wild type embryos we can see the mucus secreting cells in the external plane and the melanocytes expressing *kit* in the inner plane. In *sox10* mutants we can only detect mucus secreting cells on the external plane of the embryo, and no cells in medial plane. In *mitfa* mutant we can see the mucus secreting cells expressing *kit* in the outer plane and some bigger cells in the inner plane, which are very likely to be melanocytes.

This experiment however did not provide us a clear answer to our question. In order to get stronger evidence to support that the cells in the medial plane of *mitfa* mutant embryos are melanocytes, a double fluorescent *in situ* hybridisation was done. The aim of this experiment was to detect both *mitfa* and *kit* expression. Melanocytes express both genes, and thus we should detect co-expression. Therefore if those cells are melanocytes the signal should co-localise. In wild type we can clearly see melanocytes expressing both *kit* and *mitfa*. In *mitfa* mutants we can identify some

cells in the dorsal side of the medial plane of the embryo expressing both genes. Therefore these experiments suggest that these cells are likely to be melanocytes.

This finding meant that there is some other factor driving *kit* expression even when Mitfa is inactive in *mitfa* mutants. One strong candidate for the factor that initiates *kit* expression is *sox10*, as *kit* is not expressed when *sox10* is inactive in mutants. On these bases the initial hypothesis was modified: *sox10* might drive *kit* expression, together with *mitfa* in the feedback loop (Figure 20).

Greenhill *et al.* (2011) established that around 30 hpf most of the melanocytes showed *sox10* expression, but this decreased rapidly and by around 50 hpf signal was not detected in any of these cells. Therefore if *sox10* is supposed to drive *kit* expression in melanocyte this interaction may be transient, as *sox10* expression is restricted in time.

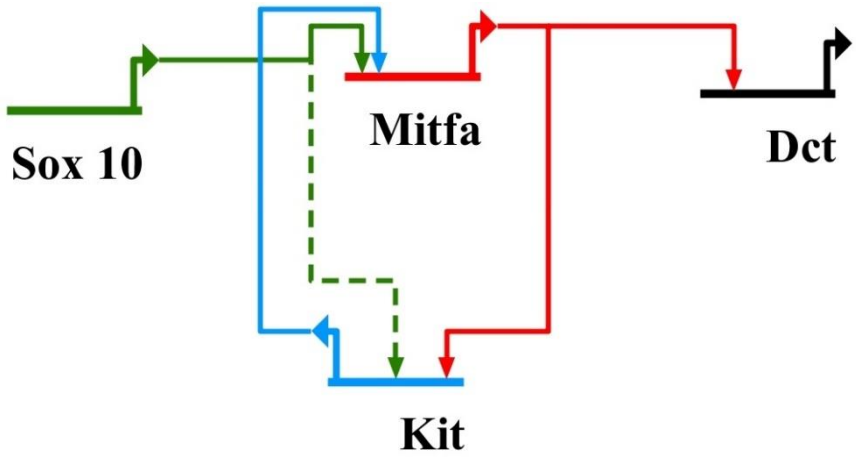

Figure 20. Revised hypothesis of the role of *kit* in the GRN. The second iteration of our hypothesis suggests that Sox10 drives *kit* expression at early stages of embryo development. *kit* is still at the centre of the feedback loop with *mitfa*, which drives *kit* expression, and *kit* itself promotes *mitfa* transcription helping to maintain its expression level to promote stable melanocyte differentiation.

kit is transiently driven by Sox10

In order to test the timeframe where *sox10* drives *kit*, we performed ISH for *kit* in wild type, *sox10* and *mitfa* mutants at three stages: 24, 30 and 48 hpf. At 24 hpf there are melanocyte expressing *kit* in *mitfa* mutant (Figure 18 C) the same as in Figure 16, but not in *sox10* mutant. By 30 hpf we cannot identify melanocytes expressing *kit* in *mitfa* mutant (Figure 18 F). This was surprising, as we could identify these cells in previous experiments at the same stage (Figure 15 and Figure 17). This might be because *sox10* expression is already decreasing at that stage (Greenhill *et al.*, 2011) and the *sox10* dependent *kit* expression is likely to end at around that stage. Also, the embryos may be in slightly different stages in the different experiments, despite being all roughly at 30 hpf, which may add to the confusion. In any case, by 48hpf there are not melanocyte precursors expressing *kit* in *mitfa* mutant (Figure 18 I), which means that the only factor driving *kit* expression by this stage is *mitfa*. Hence, the *sox10* dependent, *mitfa* independent *kit* expression seems to be limited to early stages of development and would end at approximately 30 hpf, being definitely not detectable by 48hpf. This is consistent with the *sox10* expression observations by Greenhill *et al.* (2011) where they could not detect *sox10* expression by 50 hpf.

Kit drives mitfa and maintains *its expression through a feedback loop*

The next aspect to be tested was the maintenance of *mitfa* expression by *kit* as part of the feedback loop. Mitfa expression is directly driven by *sox10* (Elworthy *et al.* 2003), but the maintenance of its expression at a high enough level to drive the melanogenic genes and drive melanocyte commitment may be achieved through the feedback loop with *kit*. Kit mutants were used to test this interaction. It has already been shown that in *mitfa* mutants, Mitfa target genes such as *dct* are absent. We also predicted that when Kit is not active, *mitfa* expression would be reduced, and that Mitfa target genes might also be reduced, although there might be a delay. The *mitfa* ISH in *kit* mutant embryos shows an expression pattern with fewer stained cells compared to the wild type embryos. The same happened with the *dct* ISH, there's a

lower number of stained cells that is even more prominent in the ventral line. One of the possible explanations for the expression pattern reduction may be because *kit* is involved in melanocyte survival and by this stage the melanocytes have already died and cannot be detected. Parichy *et al.* (1999) demonstrated that in the absence of *kit*, melanocytes undergo programmed cell death and are extruded from the skin from around 4dpf, most evident between 5 and 6 dpf. They didn't see cell death at 2 dpf and they didn't look at earlier stages. Rawls *et al.* (2003) did some elegant temperature shift experiments to address the effects of *kit* activity on melanocyte survival using a temperature sensitive allele. They reported that loss of *kit* activity at 2dpf caused melanocyte death, and therefore *kit* is important for survival from 2dpf. Despite not studying stages as early as 30 hpf, they suggest that *kit* is not required for survival at this stage.

Therefore if melanocytes have not already died by 30 hpf in *kit* mutants, the differences in expression pattern that we see in Figure 5 may be due to a defect in melanocyte migration. It has been reported that *kit* promotes the migration of melanoblasts while they disperse from the neural crest (Parichy *et al.*, 1999). Using the same temperature shifting experiment Rawls *et al* (2003) determined that *kit* is required for melanocyte migration prior to 2 dpf. This might be because the migration machinery or signally pathway may no longer be available after 2 dpf or because the extracellular matrix composition have changed by 2 dpf. However, this role of *kit* in migration may be complete as soon as 24 hpf in some melanocytes (Rawls and Johnson, 2003). In the ISH from Figure 5 there seems to be a lower number of melanocytes expressing *dct* in the ventral line, as well as a lower number of *mitfa* expressing cells in the ventral and medial line. Consequently it is likely that the *mitfa* and *dct* expression patterns are due to migration defects in *kit* mutants, as it has been previously described (Kelsh *et al.*, 2000).

However, what can we say about the *mitfa* and *dct* expression in *kit* mutants? According to our model, the expression of these genes should be reduced in each cell in the mutant embryos. From Figure 5 it seems that *mitfa* expression might be a bit weaker than in wild type, however *dct* looks similar. This is probably because the *dct*

ISH staining is saturated. In order to characterise this better we would need a quantitative measure of expression levels, which could be obtained by RT-qPCR.

Conclusion

Kit is a key gene in melanocyte development involved in migration, survival and proliferation, as it has been reported in previous studies. Therefore it seemed logical that *kit* needed to be included in our melanocyte GRN. When Greenhill *et al.* (2011) established the GRN they identified the need of a gene (Factor Y) to allow stable melanocyte differentiation. The *kit* background studies suggested that it might be a strong candidate to fulfil that role.

In this chapter we suggested that *kit* plays the role of Factor Y in the positive feedback loop in the GRN. We have confirmed that Mitfa, which is directly activated by Sox10, drives *kit* expression. Moreover, we have proposed that Sox10 also drives *kit* expression but just transiently at early stages of development until around 30hpf, while it is expressed in melanocytes. As part of the feedback loop, *kit* seems to drive and maintain *mitfa* expression. Despite not being essential for *mitfa* expression (Hou *et al.*, 2000), Kit signalling seems to promote its maximal expression to enable stable melanocyte differentiation. However this last feature needs further quantitative characterisation.

Chapter 5:

Towards a quantitative model of the GRN

Introduction

Greenhill *et al.* (2011) used a systems biology approach to identify and develop the core GRN underlying melanocyte specification and differentiation in zebrafish. They used a mathematical modelling approach in combination with experimental data in an iterative process to create a model that makes specific predictions about the properties of unidentified factors in melanocyte differentiation and explains the melanocyte development in a semi-quantitative manner.

They constructed a dynamical model of an initial GRN based on ordinary differential equations, considering the transcription levels as dynamic variables (Figure 5). The external activatory signal in the network was designed as Factor A, comprising several transcription factors whose function in the regulation of the network is poorly understood: Lef/Tcf (Wnt pathway), Sox9, FoxD3, Pax and AP2 (Greenhill *et al.*, 2011). These input factors regulate Sox10 which had been shown to directly activate the expression of Mitfa (Elworthy *et al.*, 2003), termed the master regulator of melanocyte specification, which will activate the melanogenic genes (*dct*, tyr, *tyrp1* and Si), as well as the *sox10* promoter in a positive feedback loop. Subsequently, Sox10 represses the expression of the melanogenic genes and its expression is repressed by the recruitment to the *sox10* promoter of an Hdac1 complex that deacetylates its chromatin and represses *sox10* transcription by Mitfa. In that way, Factor A-dependent *sox10* expression as well as Mitfa-dependent expression will both be inactivated. The model proposed also the existence of a Sox10-independent, Mitf-dependent transcriptional activator of *mitfa* (Factor Y) that provides a positive feedback loop to allow stable melanocyte differentiation. This feature was not

initially part of the model, but was incorporated because the model simulations highlighted the need for such a factor to maintain stable melanocyte development. Finally, Sox9b drives Sox10 and Mitfa-independent melanisation (very low level, not detectable in single mutant) and it is activated by a poorly-defined Factor B, which has a transient role and is restricted to the early phase of melanocyte development (Greenhill *et al.*, 2011).

This model allowed them to predict the evolution of gene expression of the GRN genes along the first 100 h of zebrafish embryonic development. However, the data they had was very limited: *in situ* hybridisation expression patterns that they used to get coarse gene expression data. Therefore, when they needed to find values for the parameters of the model, they chose those that produced gene expression curves that were similar in shape to what they expected to happen based on the *in situ* hybridisation data. That means that the exact quantities and details of the curves are not very rigorous and may be erroneous.

In order to develop a quantitative mathematical model to describe the GRN underlying the development of a melanocyte in the zebrafish embryo we would need to obtain gene expression measurements of the genes in the network from single melanocytes.

Obtaining single cell gene expression measurements is very challenging, and we could not do it in the lab at that moment in time. Therefore, the experimental set up that we devised was to obtain embryos from two or three zebrafish pairs and mix them together. Then, we sampled 30 embryos per stage for the 12 stages we wanted to measure: 25 of them were fixed and reserved to perform ISH for each of the four genes we measured, and the remaining 5 were used for RNA extraction and subsequent qPCR measurement. One of the challenges of measuring gene expression levels from whole embryo is that most of the genes are expressed in other tissues in the embryo besides melanocytes (mainly in the head). Also, melanocytes originate at different times within the embryo so if we want to measure melanocyte gene expression at a certain stage some of them will be at exactly that stage, but others may be younger if they are closer to the tail, as the embryo develops from anterior to posterior. Hence, to tackle this issues we decided to cut off the head and the tail

of the 5 embryos that were going to be used for qPCR measurements, keeping just the trunk (from beginning to the end of the yolk sac extension). These 5 trunks were then used to perform RNA extraction, cDNA synthesis and qPCR measurements.

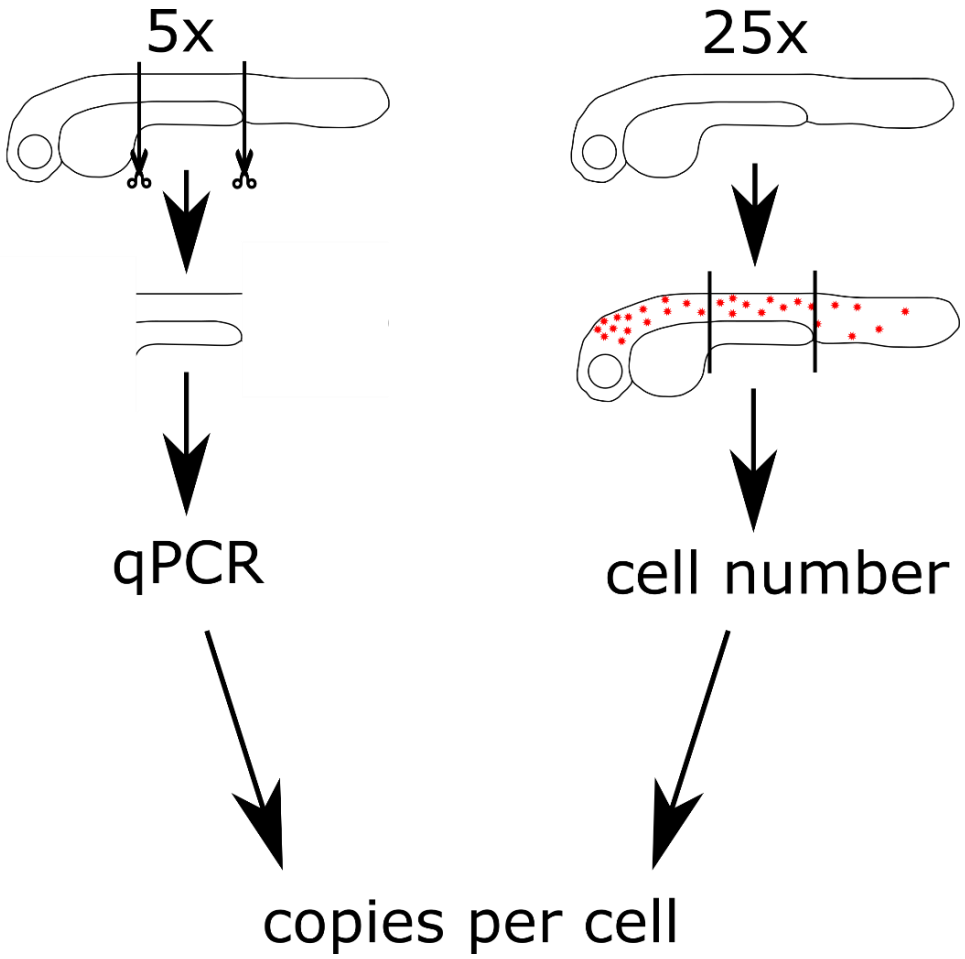

Figure 21. Diagram describing the experimental set up. Five embryos were used to cut their trunks, extract their RNA and measure the gene copy numbers through qPCR for each stage. And 25 sibling embryos were fixed at the same stage and then were used for ISH to count the number of gene expressing cells. Both measurements (gene copy number and gene cell number) were used to estimate the gene copy number per cell.

Therefore, from the qPCR data we have gene expression measurements from all the cells expressing a certain gene in the embryo trunk. To estimate the cell expression per cell we used the 25 embryos that we fixed to carry out ISH for each of the genes we are measuring. Then we counted the number of cells expressing each gene in the same region of embryo trunk that we used to obtain the qPCR measurements. Hence,

at the end we have a measurement of gene expression in an embryo trunk, and the number of cells expressing that gene in that region. So we can use the quotient of both measurements to estimate the gene expression per cell that we used to perform the parameter optimisation of the mathematical model.

All these measurements were done in 12 stages, starting from 18 hpf to 96 hpf, with the early stages more densely surveyed to capture all the gene expression variation that happens at the beginning of development: 18, 20, 22, 24, 26, 30, 32, 36, 48, 60, 72 and 96 hpf.

Results

To obtain an estimate of gene expression per melanocyte we measured the gene expression of the 4 genes we wanted to analyse (*sox10*, *mitfa*, *kit* and *dct*) in the 5 embryo trunks, and then performed four ISH for the same genes in the rest of the embryos to count the number of cells expressing that particular gene in the trunk, as described in the workflow diagram above.

To measure the gene expression through qPCR first we need to extract the RNA from the 5 embryo trunks using the TRIzol method, followed by a phenol:chloroform extraction, as described in Chapter 2. Then we need to perform a reverse transcription to produce cDNA using oligo-dT primers and finally we used that cDNA with specific primers for each of the four genes we want to measure to carry out the qPCR.

To count the number of cells expressing each of the four genes we performed *in situ* hybridisation using FastRed to label cells expressing the gene with red fluorescence. Once the ISH were performed, the cells expressing the gene in the trunk had to be counted. Initially we tried to do it using High Content Microscopy, but it was not a good method for our experimental setup. Instead we photographed the specimens using a Zeiss Axio Imager 2 microscope. We used structured illumination (Zeiss Apotome) to generate images with high resolution in each of x, y and z planes. Initially, we attempted to use an image analysis programme (MINS (Lou *et al.*, 2014)) to automate the process of quantitating the fluorescence cells. However, the

program was not designed for zebrafish embryos and our type of cells, so it did not work for us. Instead we decided to count the cells manually from the fluorescence images.

The copy number data for each gene together with the cell counts were used to estimate the copy number per cell in the embryo trunk for each stage. This data was then used in the parameter optimisation algorithm to find a set of parameters for the mathematical model that describes the evolution of the gene expression as closely as possible to the data we collected.

Copy number quantification

Polymerase Chain Reaction (PCR) is a technique for amplifying DNA that can be used when we need to amplify a tiny quantity of DNA so we get enough DNA to detect it, or when we want to compare two different samples of DNA to see which one is more abundant. If we amplify both samples at the same rate, it can be calculated which one had more copies of the target DNA to begin with. PCR is carried out by a thermostable polymerase enzyme, which synthesises a complementary strand to any single strand DNA, providing it has a double stranded starting point. This principle is what allows amplification of specific genes, as we can use primers that will hybridise with the single stranded DNA of our gene of interest to create a double stranded portion that the polymerase can use to synthesise the rest of the complementary strand. During the PCR the temperature is changed several times to control the action of the polymerase and the binding of primers. Usually the reaction starts with a rise in temperature to 95^{o}C to denature the initial double stranded DNA molecules. Then the temperature is lowered to around 60^{o}C (this will depend on the annealing temperature of the primers) to allow the primers to bind to the single stranded DNA molecule of our gene of interest. Lastly the temperature is raised to 72^{o}C as this is the optimal temperature for the polymerase. These temperature changes are repeated in cycles until we have enough amount of DNA. In every cycle the copies of our gene of interest double, so at the end of the PCR we well have 2^{n} copies, where n is the number of cycles. Then we could use the product of the PCR to quantify our

gene of interest by for example running it on an agarose gel, where be brighter the band, the more copies of the target gene there are (Saiki *et al.*, 1988).

The principle behind real-time PCR (or quantitative PCR, qPCR) is the same, but instead of looking at bands on a gel the reaction is performed in a special thermocycler that monitors the amplification in "real-time". There are many techniques to allow a PCR reaction to be monitored, but they all have in common the link of amplification of DNA and generation of fluorescence that can be detected with a camera during each PCR cycle. Hence, as the number of DNA molecules increase in the reaction, so does the fluorescence. The main benefit of this method over the normal PCR is that it allows quantitative analysis of gene expression; in comparison, normal PCR is semi-quantitative at best (Nolan *et al.*, 2006).

One of the main decisions when designing a qPCR experiment is choosing the right detection chemistry. These compounds generate fluorescence during the PCR amplification which allows the thermocycler to monitor the reaction in real time. The most commonly used chemistry (and the one we used in this project) are the intercalating dyes, in particular SYBRgreen. This dye is fluorescent on its own but in the presence of a double stranded DNA molecule, the dye intercalates with the DNA double helix, altering the dye chemical structure and increasing its fluorescence. So as the PCR generates more DNA, more dye binds to it and more fluorescence is produced. Intercalating dyes are the most economical detection chemistry method available, however they are not specific: if your PCR amplifies the wrong target, or more than one target, you will still get an amplification plot that will look like a real signal, as intercalating dyes will bind to any double stranded DNA molecule regardless of what it is (Bustin, 2000). This means that an additional analysis is required in the form of melt curve. This is carried out after the PCR reaction has finished and consists of heating the PCR products gradually from around 55^{o}C to 95^{o}C while fluorescence data is collected. At the beginning, the DNA is double stranded, bound to SYBRgreen, and therefore very fluorescent. As the PCR product is heated up the melting temperature of the PCR product will eventually be reached and the double strands will separate releasing the SYBRgreen. This results in a sudden drop in fluorescence. Subsequently all the PCR product is melted and the fluorescence

level plateaus again (Nolan *et al.*, 2006). The data is usually presented as rate of change of fluorescence. The sudden change of fluorescence is seen at the melting temperature of the PCR product is visualised as a sharp peak in a flat background when only a small change in fluorescence is occurring. So in the perfect reaction where a single PCR product is amplified a single peak should be visualised. The melting temperature of our target gene amplicon can be calculated and therefore we can use the melting curve to check that we have amplified the right product. If more than one peak is observed it may indicate issues in the reaction such as primer dimers or amplification of genomic DNA.

One of the most used ways to analyse gene expression is by relative quantification, which means assessing how the gene expression levels change between one sample relative to another. This is simpler to perform than absolute quantification and sometimes provides all the information we are looking for. This method is based on the doubling of the target DNA amount in every cycle of an efficient PCR reaction. If a PCR reaction reaches the exponential growth one cycle earlier than another reaction, then the former contained twice the target DNA at the beginning of the reaction (Wong and Medrano, 2005). However, for some experiments we need to know precisely the number of copies of a target DNA molecule in our sample. For example, when using qPCR for as a diagnostic tool to measure HIV viral load in the patient in order to know the dose of drug to be administer. In our case we need to know the exact copies to perform the parameter optimisation procedure in our mathematical model. To perform absolute quantification we need to use standard curves using a sample that contains our target gene at a known concertation. To achieve this we can clone our target amplicon into a vector (in our case a plasmid). The precise concentration of this vector is then measured to know the exact number of copies of the amplicon and then a serial dilution is performed (typically a 10 fold serial dilution). The user measured the Ct or Cp values (the cycle number at which the fluorescence signal crosses the threshold into the exponential growth) of the standard curve points and then compare them to the Ct values of the unknown sample to read off the number of copies of the target gene present in the sample (Bustin, 2000).

For the standard curve we used plasmids containing the gene of interest cDNA that were initially designed and used to produce ISH probes. The gene insert, plasmid backbone and total length of the plasmid can be found in Table 3 in Chapter 2. These plasmids have been tested many times and they produce a probe that hybridises successfully with the mRNA in ISH. We used our primers to amplify our amplicon in a conventional PCR and all the four plasmids contained the right amplicon of approximately 100 bp. Therefore we used these plasmids to produce the 10-fold dilution standard curves. Initially we produced a 7 points curve, but when we tested it against our sample we realised that it was out of range in the bottom part of the curve, so we removed the top 4 points of the curves and added two more points at the bottom end of the curve, so our samples are in range. The standard curve results are usually plotted in a graph with Ct values versus the log of the cDNA quantities and a linear regression is obtained to fit these values. The slope of the linear regression equation measures the assay's efficiency, with a slope of -3.32 indicating 100% efficiency. This is because it takes 3.32 cycles in a 100% efficient qPCR amplification to increase the number of molecules 10-fold in each dilution ($2^{3.32}=10$) (Nolan *et al.*, 2006). Figure 22 shows the standard curve for *dct* including the sample data points, as well as the linear regression equation and efficiency and R^2 data. As it can be seen our unknown samples fit in the standard curve range and the efficiency is very close to 100%. Also R^2 that measures the performance of the assay is very high, which indicates that the data fits very well the standard curve linear regression used to calculate copy numbers. The standard curves for the other three genes can be found in Appendix 5. The efficiencies of our assay are all within the acceptable PCR efficiency of 90-110% (or very close in the case of *mitfa*). The R^2 of our assays are all higher than 0.98, which is the indicative of a stable and reliable assay (Nolan *et al* 2006).

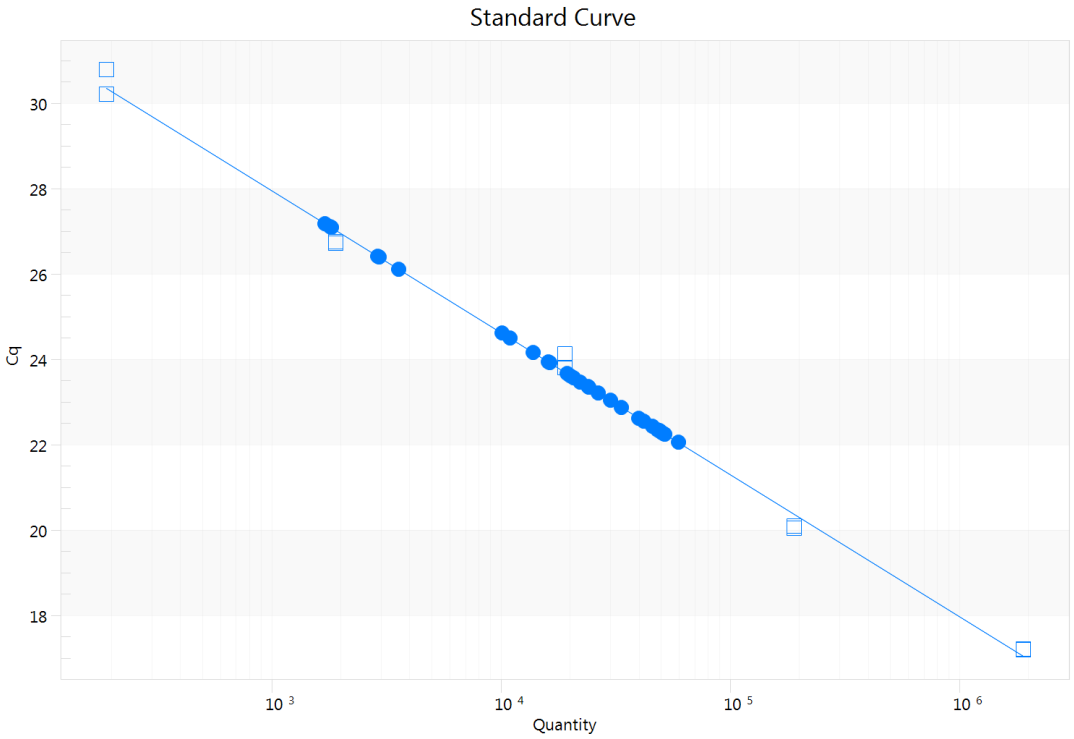

Figure 22. *dct* standard curve. The squares represent the five points of the standard curve by duplicate. The solid circles represent the unknown sample data points that are being measured. The linear regression equation is: y=-3.3x+37.84, R^2=0-996, Efficiency=99.73%.

A qPCR plate was run for each of the four genes. The samples were run in triplicate, and the standard curve points were run in duplicate. A no-template control (NTC) was also run, which contained all the reaction elements but the cDNA template. The NTC was negative in all of the plates, and the melt curve analysis carried out after the qPCR amplification showed a single sharp peak, indicating that only our target amplicon was amplified and that complications due to primer dimers or genomic DNA amplification were not an issue.

The results of the qPCR can be seen in the graphs below. The bars represent the mean copy number of the three replicates per stage, and the error bars represent the standard deviation.

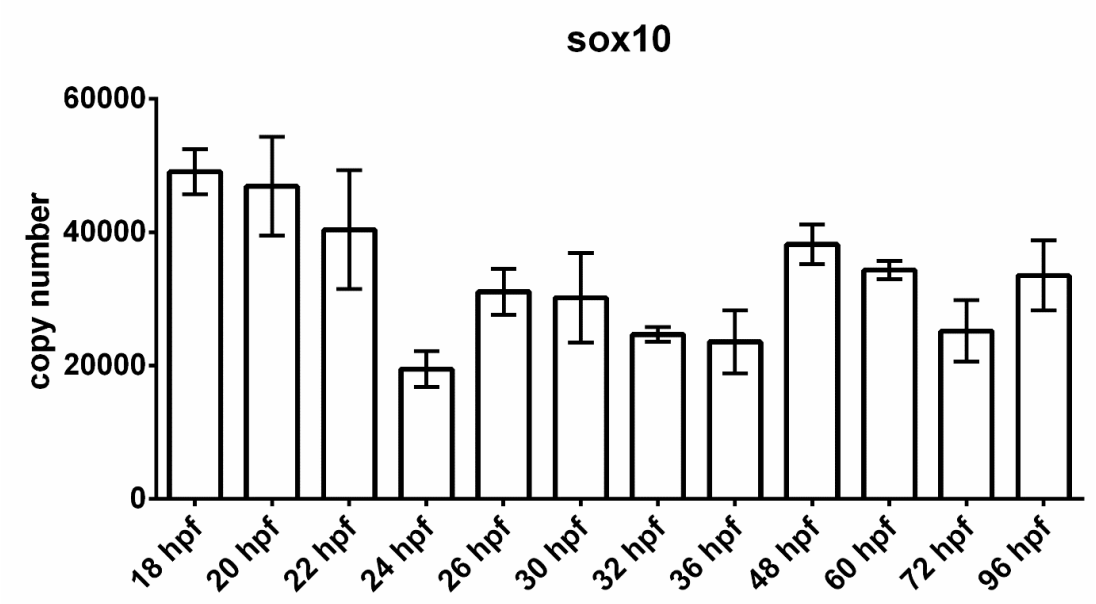

Figure 23. *sox10* qPCR copy number data. Each of the columns represent the mean of the three measurements per stage and the error bars represent the associated standard deviation.

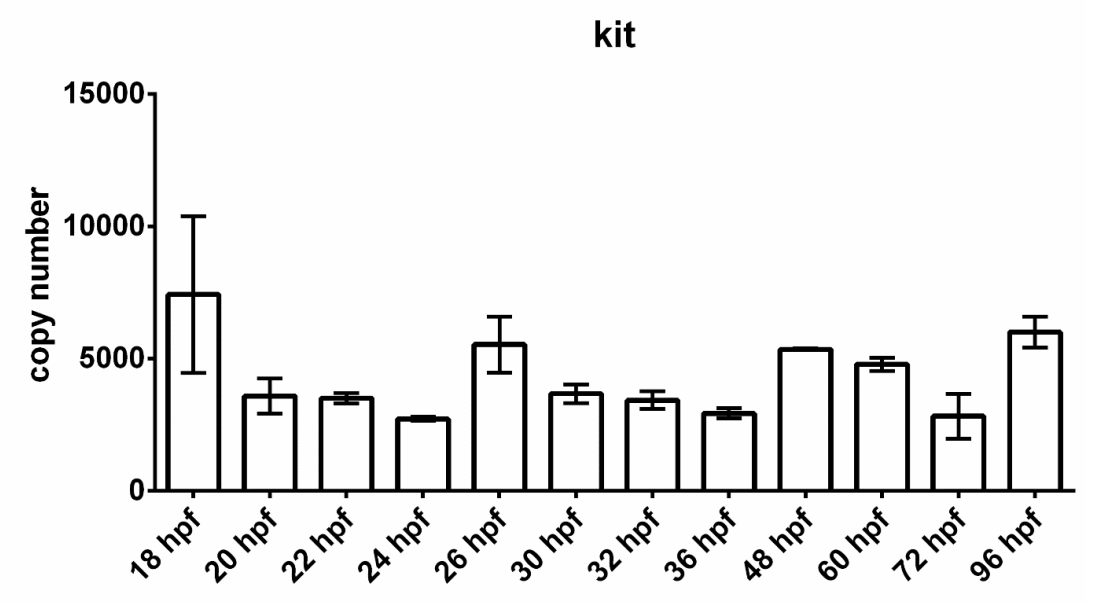

Figure 24. *kit* qPCR copy number data. Each of the columns represent the mean of the three measurements per stage and the error bars represent the associated standard deviation.

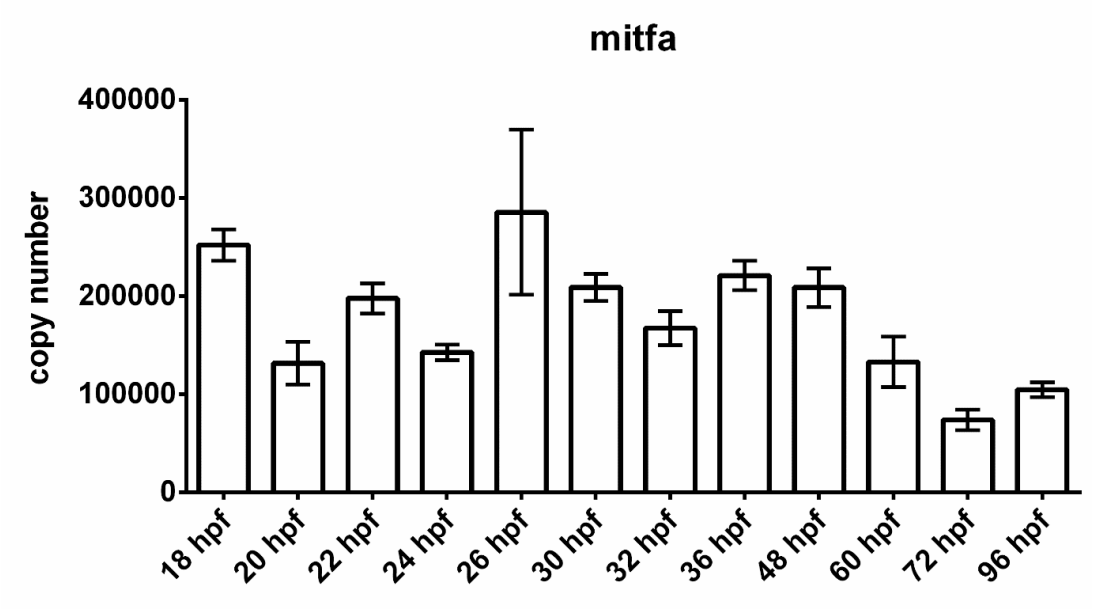

Figure 25. *mitfa* qPCR copy number data. Each of the columns represent the mean of the three measurements per stage and the error bars represent the associated standard deviation.

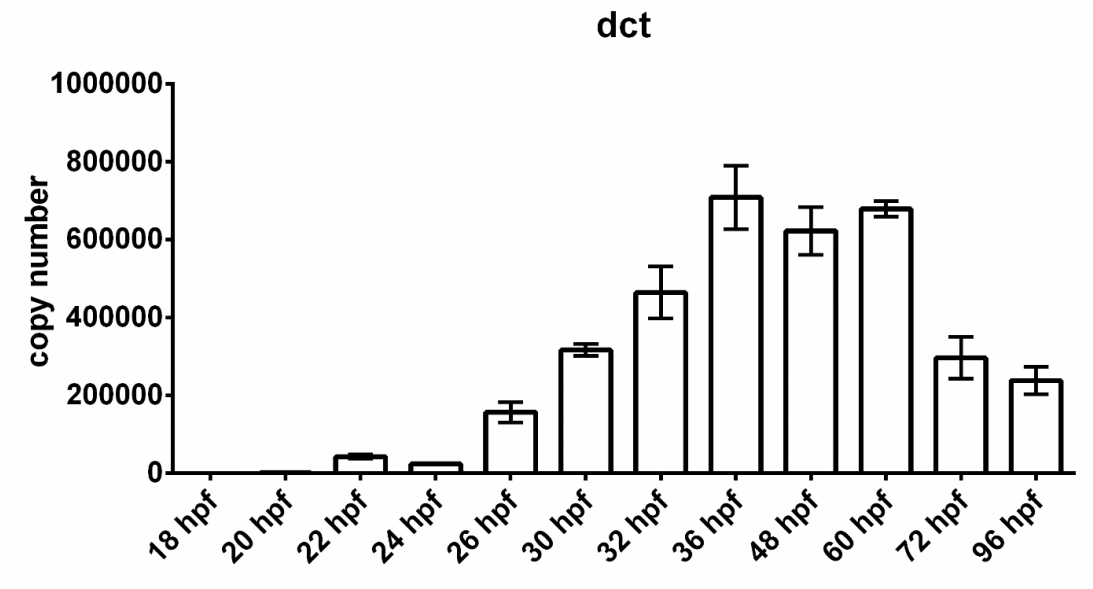

Figure 26. *dct* qPCR copy number data. Each of the columns represent the mean of the three measurements per stage and the error bars represent the associated standard deviation.

As in many biological experiments our qPCR results are likely to be influenced by noise. Noise could be generated by organism-to-organism variation in mRNA levels, marginally different staging of the embryos within the same time point, or technical noise generated by slightly different measurements from the thermocycler

(Bengtsson *et al.*, 2008; Brennecke *et al.*, 2013). So it may be useful to try to reduce the noise by a process called smoothing. In smoothing, the data points of a signal are modified so that the points that are higher than the immediately adjacent points (presumably because of noise) are reduced, and points that are lower than the adjacent points are increased. This leads to a smoother signal with reduced noise without distorting the true signal too much (Spiess *et al.*, 2015).

The typical smoothing method is based on a moving average filter which smooths the data by replacing each data point with the average of the neighbouring data points. The method we used is called Savitzky-Golay filtering, and it is a generalised moving average with filter coefficients determined by an unweighted linear least-squares regression and a polynomial model of given degree (Matlab technical documentation).

The smoothing procedure was carried out in Matlab (Appendix 6) and it started by loading the copy number data and associate standard deviation produced by the qPCR software. Then the copy number data was smoothed with the function “smooth” and the method “sgolay” (Savitzky-Golay). The smoothed, as well as the original data for each gene are shown in Figure 27.

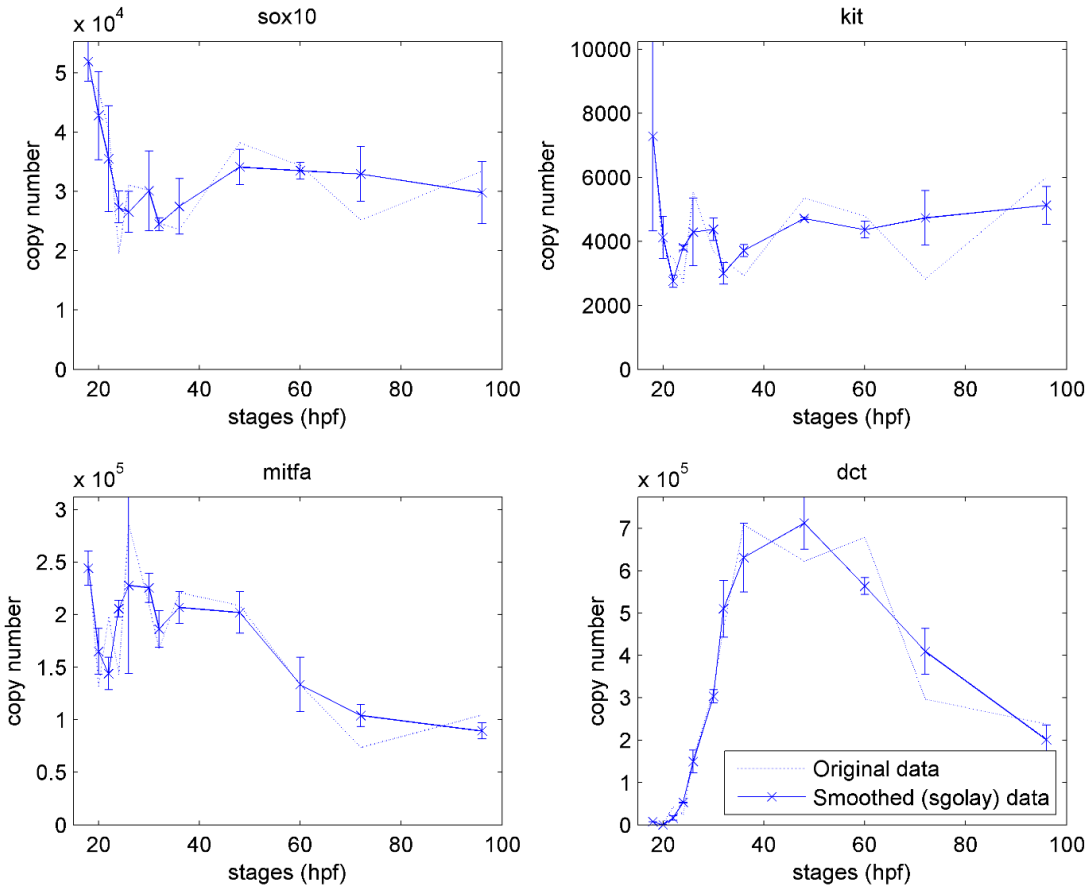

Figure 27. Smoothed copy number data from *sox10*, *kit*, *mitfa* and *dct* produced in Matlab using the Savitzky-Golay algorithm. The plots show the original data in the dotted line and the smoothed data in solid line with its associate standard deviation.

As we can see in Figure 27 the smoothing procedure has reduced the big jumps in gene expression in between contiguous stages, making the changes in gene copy number more even in between closer stages, which it is what we expect to find in the organism: the copy number changes progressively rather than though big jumps. This data will be used in conjunction with the cell number counts to estimate the copy number per cell for each of the four genes.

Cell number count

We have obtained the copy number measurement of the four genes we are investigating through qPCR. To estimate the copy number per cell for the four genes we need to know how many cells are expressing our genes of interest in the trunk at

each of the stages we have measured. For this we carried out ISH for the four genes at the 12 stages selected using FastRed as a dye. Once the ISH were done we could start to count the labelled cells in the trunk region (same region we used to extract the RNA and perform the qPCR). Our aim was to develop an automated method to take the images and analyse them using a software to count the cells. That is why we investigated the use of high content microscopy (HCM), which allows the automation of image acquisition and its software has the capability of counting cells from the images acquired. Unexpected difficulties resulted in us adopting a manual imaging approach, using a Zeiss Axio Imager 2 with structured illumination attachment (Zeiss Apotome technology) to manually take Z-stack images of our ISH specimens. We initially attempted to implement an algorithm to perform automated cell counting. Regrettably this did not work as expected either. So finally we decided to count the relevant cells from the Apotome images manually.

HCM

High-content microscopy (HCM) is a powerful technique for high-throughput microscopic analysis of cells or tissues. HCM involves probing the cells or tissues, in a plate or on microscope slides, with single or multiple fluorescent dyes; capturing images of the cells very rapidly with a high-resolution imaging instrument; and then extracting detailed information from the images with powerful software. The automation and throughput provided by HCM allows you to get a lot of information and very quickly in comparison to normal microscopy.

The high content microscope we used is the IN Cell Analyzer 2000. This microscope reads plates up to 96 wells, but it can also image microscope slides by using an adaptor. The plate (or slide adaptor) is placed at the top of the instrument and then it takes it inside to acquire the images. All the settings of the microscope are controlled from its software in the workstation. Once all your samples are in the plate wells and you have loaded the plate into the microscope, the next step involves the setup of the image acquisition parameters: wavelengths to record, objective to use, exposure, and focal plane to image. This last parameter is quite difficult to adjust. If your sample is a flat two dimensional tissue that sits on the bottom of the well surface imaging is relatively simple, but if you sample is a 3D tissue or organism you need to

tell the software an offset parameter that measures how far from the bottom of the well your plane of interest is. This microscope seems to be designed for cell culture or very thin tissues that sit on or very close to the bottom of the plate wells. The main issue we found when trying to image zebrafish embryo was that we needed to position them manually at the bottom of the well and on their side, which is the orientation we needed to image to obtain the best cell counts. Despite doing this, the embryos will sit in slightly different focal planes in each well so the focal offset parameter will be different in each well. All these issues made impossible for us to automate the image acquisition of ISH zebrafish embryos. Therefore we switched to manual mounting and individually imaging each embryo in the well of the plate.

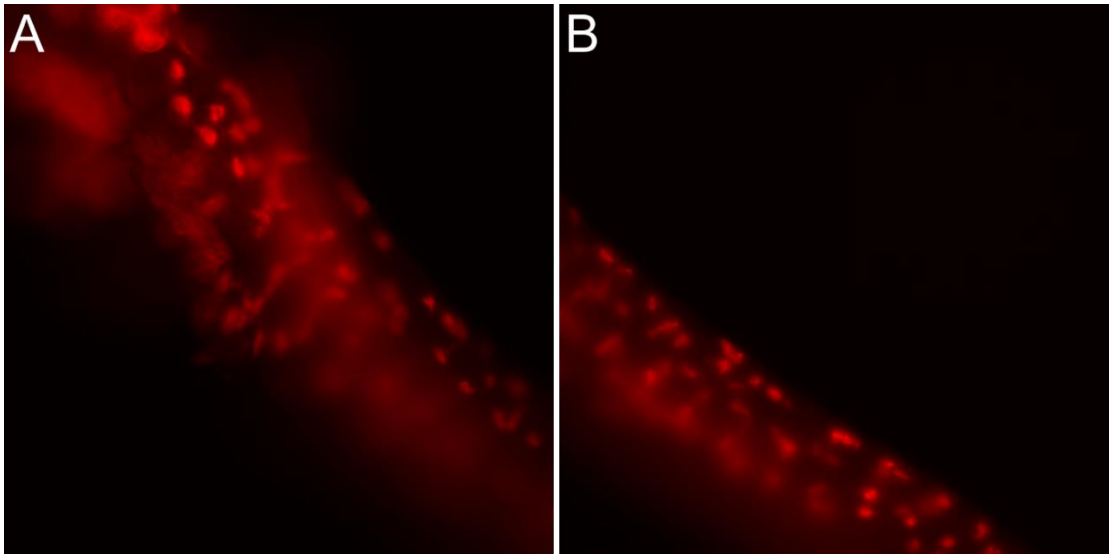

Figure 28. 30 hpf ISH embryos for *dct* using FastRed imaged by the IN Cell Analyser 2000 HCM. These two images are from the same embryo and one single focal plane. Image A captures a region of the embryo closer to the head. Image B captures a region of the embryo closer to the tail. Image B is a continuation of A. Image orientation: embryo head is on the left and tail on the right of the images.

Once the settings were adjusted, the microscope acquires the images that then can be analysed with the system's software, which has tools to segment and count cells. The software works really well with cell cultures and thin tissue layers, but zebrafish embryos are very thick and the melanocytes stained are in different planes within the embryo. So when an image is captured from a focal plane we also get scattered fluorescence coming from the stained cells in the surrounding planes of the embryo,

decreasing the signal to background ratio (Figure 28). This makes very difficult for the software to identify and count the stained cells. Due to these issues with the HCM, we decided that this was not the right method to count cells in the zebrafish embryos for our experiment. Instead we decided that mounting and imaging the specimens in a conventional microscope, and then using software to identify and count the cells would be a better solution for us.

Apotome and MINS software

We have in the lab a state of the art Zeiss Axio Imager 2 with the Apotome.2 technology. The Axio Imager 2 is a fluorescence microscope with outstanding optics, but what makes it very useful for our experiment is the structured illumination Apotome module integrated with the microscope. This technology allows you to perform optical sectioning in fluorescence images. Apotome projects a grid structure into the focal plane of your specimen, then moves it into three positions using a scanning mechanism. At each grid position, the system acquires an image. The three images are then processed to produce a resulting image of one optical section with improved contrast and increased resolution. It is especially useful to prevent scattered light from out of focus planes in thick specimens (such as embryos), producing images with high contrast and resolution, and very low background. It also has the advantage that the system remains as easy to operate as a conventional microscope (Zeiss technical documentation).

We used this technology to take Z-stack Apotome images of our specimens. To take the best images with the microscope we needed the specimen on lateral position and as flat as possible, ideally just with a coverslip on top, without using spacers in between the slide and the coverslip. In this way we would reduce the number of planes to image on the Z-stack and will make sure that we have many cells in focus on each plane. When we initially tried to put a coverslip on top of the glycerol drop containing the embryo, the head and yolk sac were squashed and the yolk sac burst releasing a lot of debris that would hinder a good photography. Therefore to avoid this we dissected the head and the yolk sac of each embryo, keeping the trunk, which is the part we are interested in and we need to count, and the tail. This made the specimens very flat and allowed us to use just one coverslip on top of the embryo.

3D images of four individual embryos were taken for each stage. These 3D images were then used to count the cell number.

The software that we selected to perform this task was MINS (Modular Interactive Nuclei Segmentation) software developed by Xinghua Lou (Lou *et al.*, 2014), and recommended by our colleague Silvia Muñoz Descalzo, as she used it successfully to count and analyse stem cell cultures and preimplantation stage mouse embryos (Le Bin *et al.*, 2014). MINS is a program that runs in Matlab and was designed for nuclei segmentation. The software consists of three modules: detection, segmentation, and classification. The detection parts locates the cell nuclei. Then by segmentation, the detection output is expanded to cover the full nuclear body. And finally the classification module removes outliers, gives a result table with a quantitative analysis of fluorescently labelled cells, and allows classification of the cells depending on their position in the image (Lou *et al.*, 2014). The software was designed to analyse confocal Z-stack images, so it worked well with our Apotome 3D files. However, MINS was programmed to identify nuclei, and therefore it struggled to recognise the red labelled melanocytes, as their shape is very different to that of nuclei.

We then asked whether staining of the melanocyte nuclei (Hoechst stain, post ISH; nuclei blue) would allow the software to identify them and then allow the expansion of the nuclei to include the entire cell. Initially we tried it in just a batch of embryos from *mitfa* ISH to test the method. The idea was that first the nuclei would be detected using the blue channel image, then the segmentation tool would assess the red channel data, expanding the nuclear area in those cases where it found red fluorescence surrounding the nucleus to cover the whole melanocyte, and finally those cells with red fluorescence were counted as melanocytes. However, zebrafish embryos have thousands of cells in the trunk, and when using the detecting module it needed a lot of time and a huge computer power, which was not available, therefore it crashed every time. We arranged the use of a high performance computer to do this analysis but first we needed to probe that it worked. So we extracted a subset of 4 planes from the original Z-stack to try to reduce the number of nuclei on the images to be identified. We run the detection module for over a weekend and we managed to obtain a file with the nuclei identified (Figure 29), but

when we tried to segment the melanocytes by expanding the nuclear signal the program detected way more melanocytes that there actually were on the image, as it assumed some of the background (even as low as it was) to be real red signal coming from the melanocytes. Also, these cells have an irregular shape which make them a quite challenging to segment. Moreover, the segmentation took several hours again. After all these testing we concluded that this method was not useful to perform the cell count in our experimental setup.

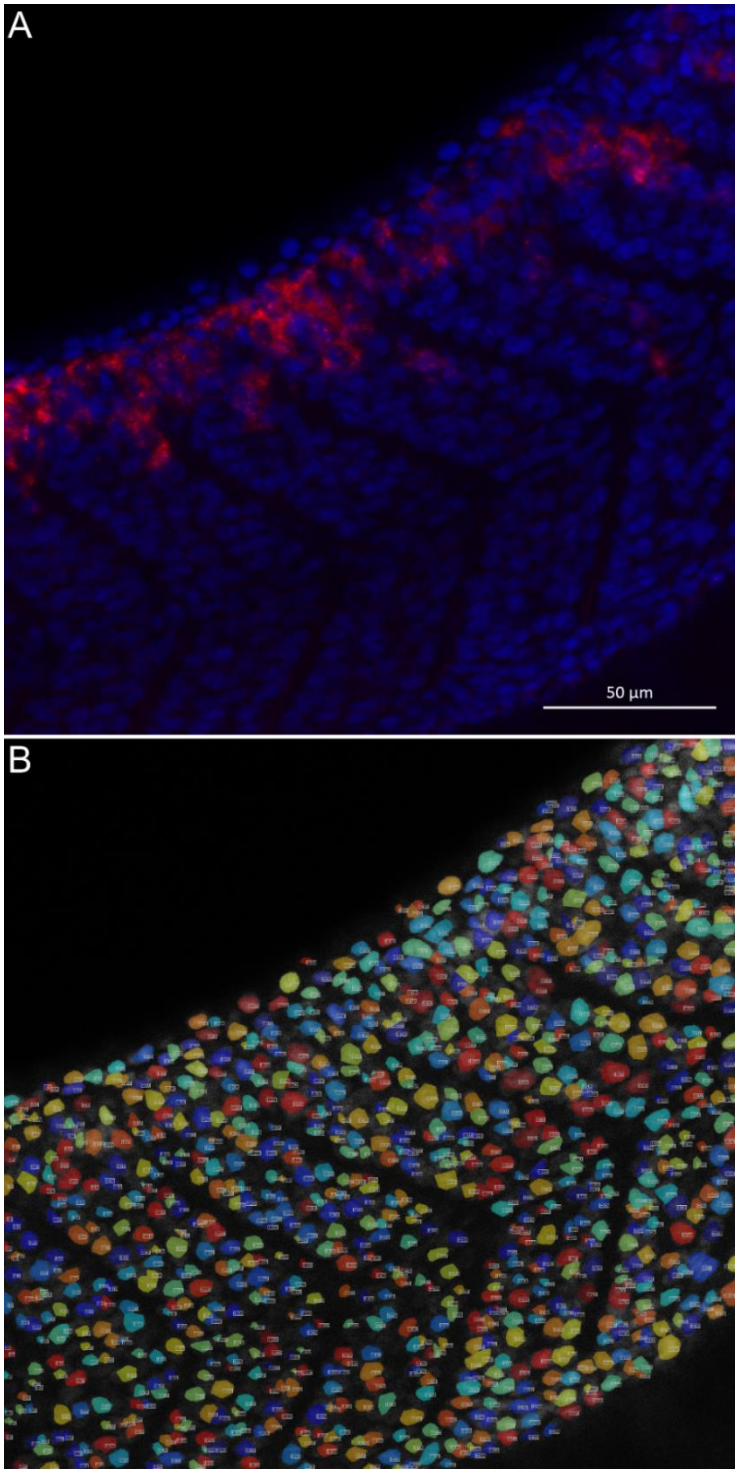

Figure 29. Image from a *mitfa* ISH stained with Hoechst to label the nuclei. The image was taken using a 63x objective, and a subset of 4 planes out the whole Z-stack was used to detect the nuclei using MINS (Panel B), which managed to identify most of the nuclei quite successfully. Images taken from the mid trunk, above the yolk sac extension. Image orientation: embryo head is on the right and tail on the left of the images. Scale bar for both images: 50 µm.

We came to the conclusion that the fastest and more accurate way to count the cells was by doing it manually by eye using the Apotome 3D images and the software ImageJ.

Microscopy

We acquired four Z-stacks of the trunk region per stage using the Apotome. We took images of the 12 stages processed for the four genes.

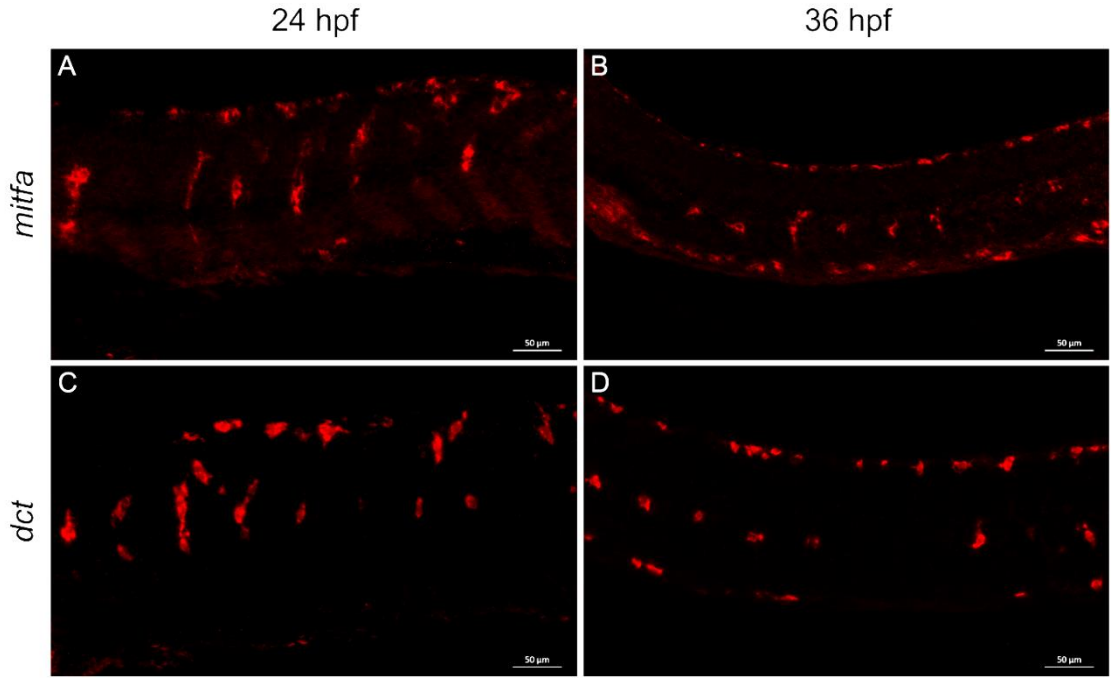

Figure 30. This panel shows an example of single plane of the Z-stacks imaged using the Apotome from *mitfa* and *dct* ISH at stages 24 and 36 hpf. The Apotome technology produces high contrast and high resolution images where cells are quite easy to count manually. Images taken from the mid trunk, above the yolk sac extension. Image orientation: embryo head is on the left and tail on the right of the images. Scale bar for all images: 50 µm.

Once the images were collected we proceeded to count the cells. For that we used a cell counter plug-in in ImageJ that allows you to mark each cell you have counted

with the mouse. It also keeps those markers in the different planes of the Z-stack to avoid double counting of the same cell present in adjacent planes. Once you have gone through all the planes in a stack, the plug-in gives you the number of cells you have counted.

First of all we counted the *dct* ISH, as this gene is a melanocyte differentiation marker and provides the better estimate of melanocytes and their precursors at each stage (Elworthy *et al.*, 2003). We imaged four specimens per stage and we counted each image three times, producing cell counts that can be seen in Figure 31. However, as *dct* is a marker for melanocyte differentiation it is not the best gene to estimate the cell number at early stages development, when the cell is differentiation. For that the best estimate comes from the *mitfa* ISH (Montero-Balaguer *et al.*, 2006; Nguyen *et al.*, 2010). Nevertheless, not all the neural crest that turn on *mitfa* will become melanocytes. Some will become iridophores instead, as there is a bipotent melanophore/iridophore precursor that expresses *mitfa*. By 32 hpf, *mitfa* positive bipotent precursors will have chosen to become melanocytes or iridophores. Hence, we predicted that at some point, likely around 32 hpf, the two markers of melanocyte lineage cells would coincide, at a stage when all cells were expressing both markers at detectable levels (Curran *et al.*, 2010). We counted *mitfa* expressing cells until the stage when both *mitfa* and *dct* numbers match, which happened at 32 and 36 hpf, as can be seen in Figure 32. For *sox10* and *kit* we counted only half of the stages, as we soon realised that these were not going to be used to calculate the copy number per cell, since both *sox10* and *kit* are expressed in many tissues other than melanocyte lineage cells. And hence these genes would be a bad estimate of melanocyte number.

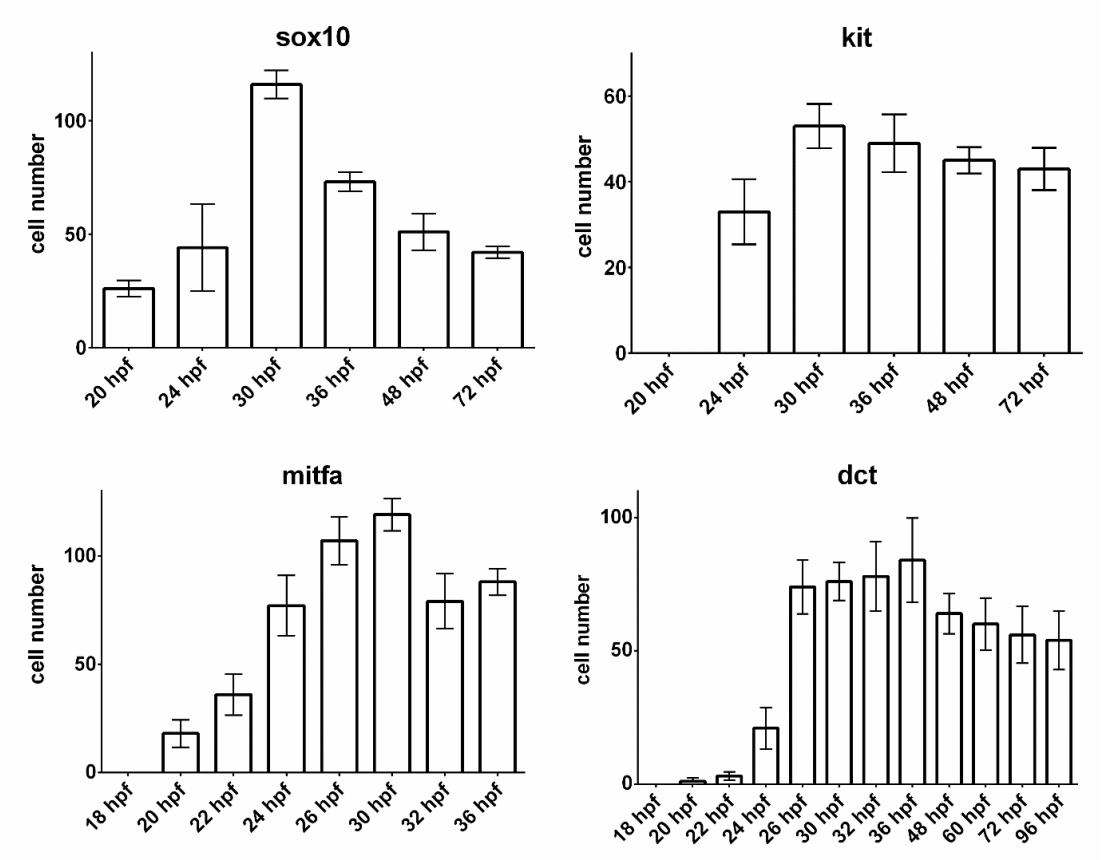

Figure 31. The plots above represent the mean cell number count per stage in the four different genes that we are investigating. The plots show also the associated standard deviation error bars for each stage.

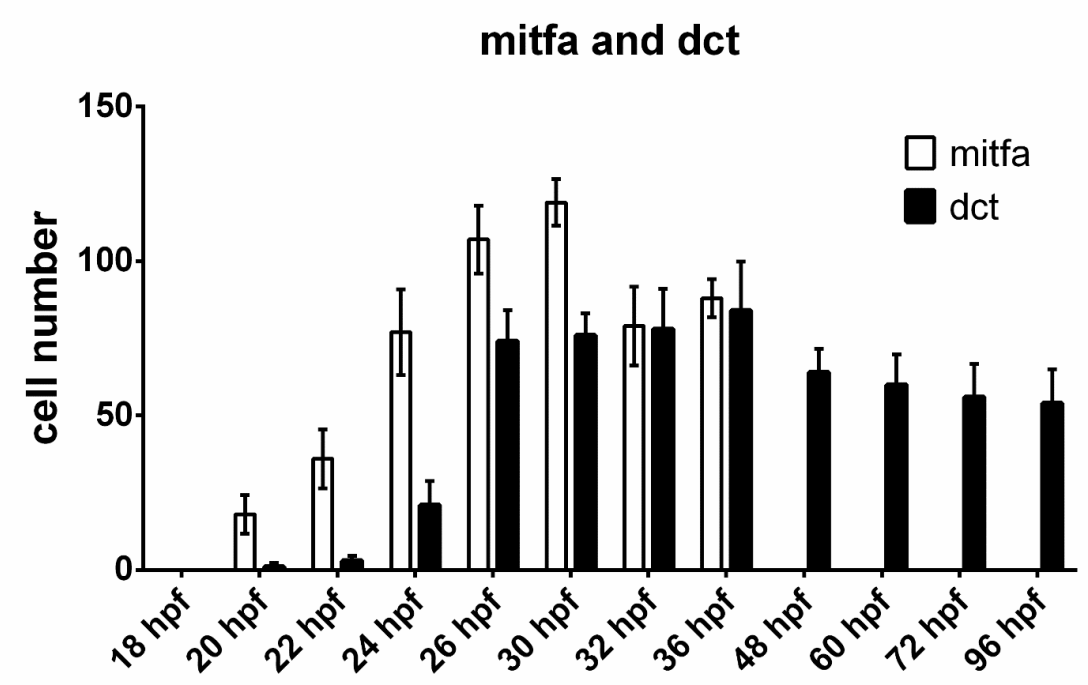

Figure 32. *mitfa* and *dct* cell counts. The number of cells match at 32 and 36 hpf, Therefore *mitfa* cell counts were used until 32 hpf and the *dct* cell counts were used for the rest of the stages.

In order to be sure that the melanocyte cell counts from the ISH represent the real melanocyte number, we compared the *dct* ISH counts at 60, 72 and 96 hpf with counts of melanised melanocytes in wild type embryos that did not go through an ISH experiment. The melanised melanocyte count (Figure 33) numbers for the three stages are very similar to the numbers of the *dct* ISH for those same stages, confirming that *dct* expressing cell counts represent a good estimate of real melanocyte numbers.

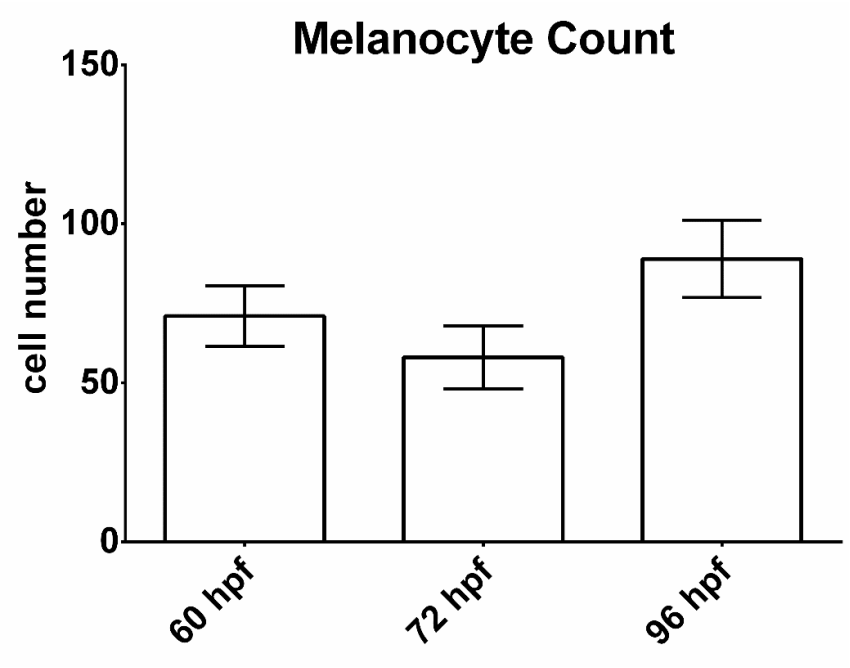

Figure 33. Melanised melanocyte mean counts from untreated zebrafish embryos with standard deviation error bars

As with the qPCR copy number data we performed a smoothing procedure with the cell number data. We used the cell count numbers from the *mitfa* ISH for the early stages of development, as this gene is a better marker of melanocyte specification than *dct*, which is a better marker of differentiated melanocytes (Lister *et al.*, 1999; Elworthy *et al.*, 2003). Therefore we used *mitfa* cell numbers until they match the number of *dct* expressing cells (32-36 hpf); from that point onwards we use the *dct* cell count.

The data smoothing process was carried out using Matlab and the same method as before (Savitzky-Golay). The smoothed data (Figure 34) will be used with the smoothed copy number data to produce an estimation of copy number per cell.

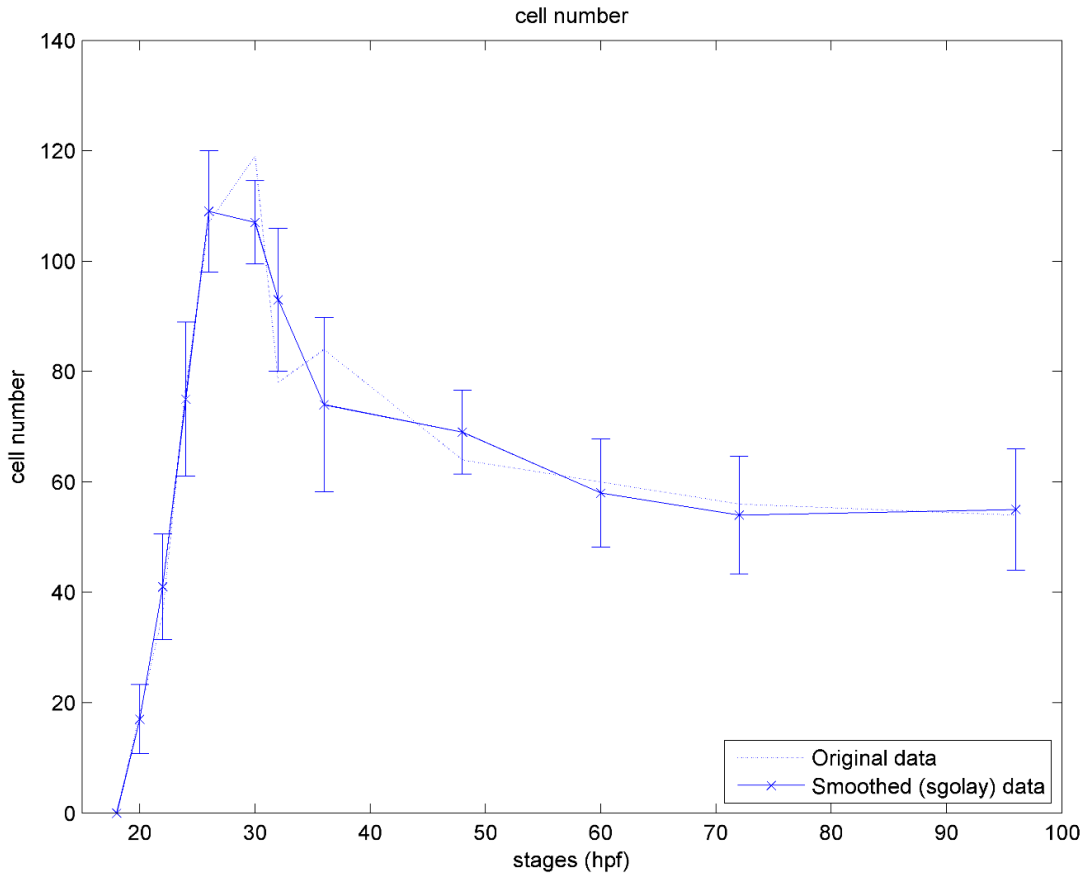

Figure 34. This graph shows the cell count smooth data in the solid line and the original data in the dotted line. This cell number counts come from the *mitfa* counts from 18 to 32 hpf and the *dct* counts from then onwards.

Copy number per cell estimation

The copy number measured from the four genes through qPCR together with the cell number estimated from *mitfa* and *dct* cell counts were going to be used to estimate the copy number per cell that we needed to perform the parameter optimisation of the mathematical model to make it quantitative.

Sox10 is expressed in many tissues other than melanocyte lineage cells: glial cells, enteric and sensory neurons, dorsal root ganglia, oligodendrocytes, among others (ZFIN). Kit is also expressed in a wide variety of tissues besides melanocytes: epiphysis, gut, heart, mucus secreting cells, notochord, optic vesicle, proctodeum, among others (ZFIN). Therefore the copy numbers that we measured though qPCR

come from all those tissues, not only melanocytes. The same happened with the cell numbers we counted: only a portion of the *sox10* or *kit* expressing cells are melanocytes, the rest would be other cells expressing those same genes. Because of this, we concluded that the copy number per cell calculated using the copy number and cell number measurements for both *sox10* and *kit* would be very far away from a real estimate, and hence we decided not to use produce this estimate and not to use it for the parameter optimisation procedure.

Therefore, we used the smoothed data for both copy numbers and cell numbers for *mitfa* and *dct* to estimate the copy numbers per cell in the zebrafish embryo trunk.

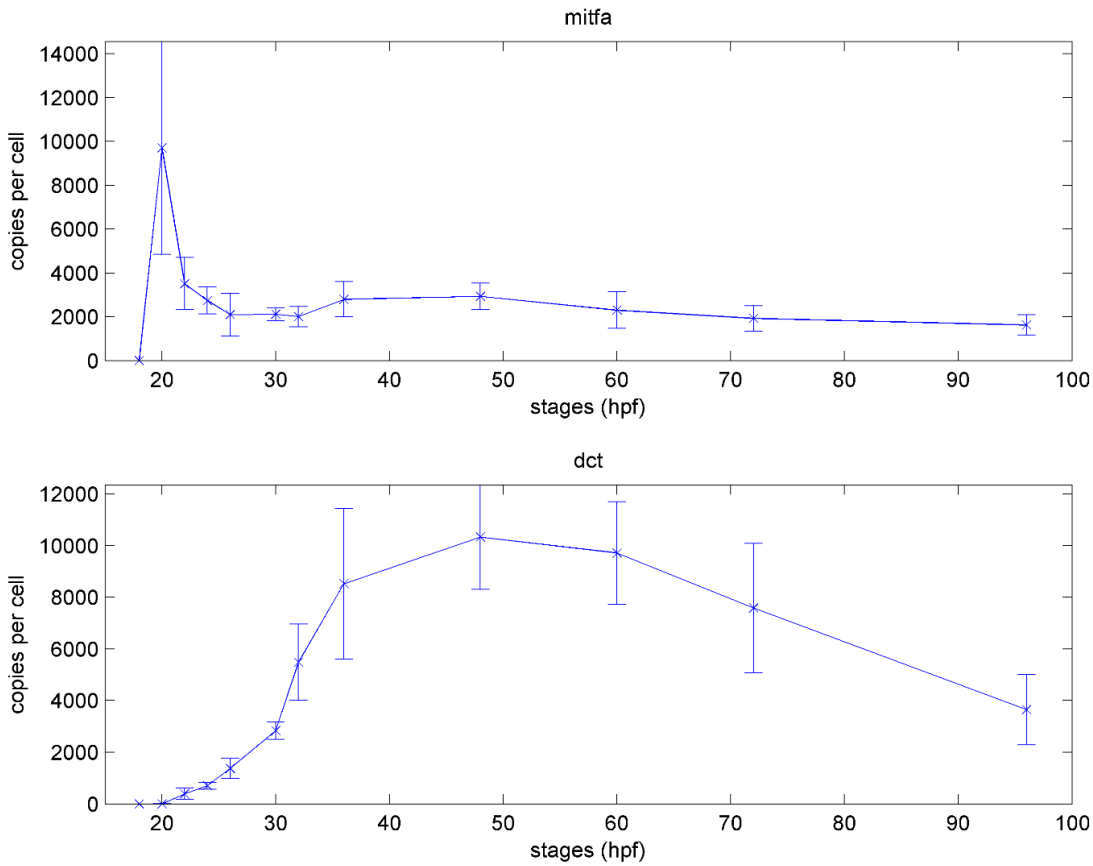

Figure 35. Copy number per cell in the embryo trunk calculated dividing copy number by the cell number for each gene. The plots show the standard deviation error bars.

For *mitfa* we saw no detectable expression in melanoblasts at 18 hpf (Figure 31). It was therefore surprising to see quite a high copy number for *mitfa* expression at 18

hpf. When the quotient of copy numbers and cell numbers was calculated the result for 18hpf was undefined, as we had to divide by 0. In this instance the Matlab script produced a value of 0 copies per cell at 18 hpf. For *dct* both copy number and cell number were 0 at 18 hpf. Hence we decided to remove the 18hpf time point from our analysis as this estimate was probably erroneous and would be a confounding factor in the parameter optimisation.

Parameter optimisation

The parameter optimisation aims to find the best parameter values so the mathematical model predictions are as close as possible to the experimental data collected. For that we use an objective function that calculates the error from the predicted trajectory calculated with a set of parameters and the measured data; and we try to minimise that error finding new parameter values through a nonlinear least-squares solver.

The error function calculates the difference between the copy number value from the data (Y_{meas}) produced by qPCR and the copy number value predicted by the differential equation solver (ode45) with a set of parameter values (Y_{pred}). This is then divided by the standard deviation of the measured data (σ_{meas}) as a way to penalise those predicted values that are too far from the measured data point:

$$E = \frac{Y_{meas} - Y_{pred}}{\sigma_{meas}} \qquad (1)$$

This is calculated for every time point of the data values. Then the objective function returns a vector (called lsqerr in our Matlab code) for the 'squares' to be minimised by the nonlinear least-squares optimisation algorithm (lsqnonlin).

In the error function Y_{meas} represents the copy number per cells value that was calculated by the quotient of the copy number and the cell number for each gene. However, to calculate the σ_{meas} we cannot use the quotient of σ from the copy number and the σ from the cell number. Instead it is calculated by:

$$\Delta C = \frac{\bar{Y}\Delta X + \bar{X}\Delta Y}{\bar{Y}^2} \qquad (2)$$

Where ΔC represents the copies per cell standard deviation, $\bar{Y}$ is the cell number, ΔX is the copy number standard deviation, $\bar{X}$ is the copy number, and ΔY is the cell number standard deviation; this is calculated for each stage.

The nonlinear least-squares optimisation algorithm starts looking for a parameter set that minimised the objective function using an initial value of the parameter set. Then it will modify the values of the parameters until it finds a minimum for the objective function. After that, our Matlab script will calculate again the mathematical model trajectories using the ODE solver and will plot them in a graph together with the data points to be evaluated.

This is repeated several times using as initial parameter values the output values from the previous optimisation round. We can also adjust the error function so the optimised parameters produce solutions that are as close as possible to the data.

The first time that we ran the optimisation algorithm (Appendix 8) we used the parameter set from Greenhill *et al.* (2011) as initial parameter values to be optimised. In this list of 59 parameters the first 8 correspond to the value of the initial concentration of each of the genes of the network at time 0, which was set at 0 (Appendix 1). We selected parameters from 9 to 59 to be optimised through the nonlinear least-squares optimisation algorithm using this objective function lsqerr=(Ymeas-Ypred)./(YmeasStD) (same as in equation 1).

Figure 36 represents the trajectories of the mathematical model from Greenhill *et al* (2011) using their parameter values. As it can be seen the units in the Y axis of the graph are in the order of ten or less, as the parameter values. However, our copies per cell data for *mitfa* and *dct* are in the order of ten thousand (Figure 15). If we perform the parameter optimisation using the Greenhill parameters as initial values the algorithm will struggle to fit the ODE trajectories to the data as it will be several orders of magnitude higher. Hence, we re-scaled the data so it was in similar dimensions to the parameters, under which conditions the parameter optimisation algorithm is able to find a parameter set that minimises the objective function. The factor that used to re-scale the data is (2/3000), which will bring the copies per cell

numbers to around ten or less. This re-scaling is reverted after the parameter optimisation is done, allowing estimates of copies per cell to be presented accurately.

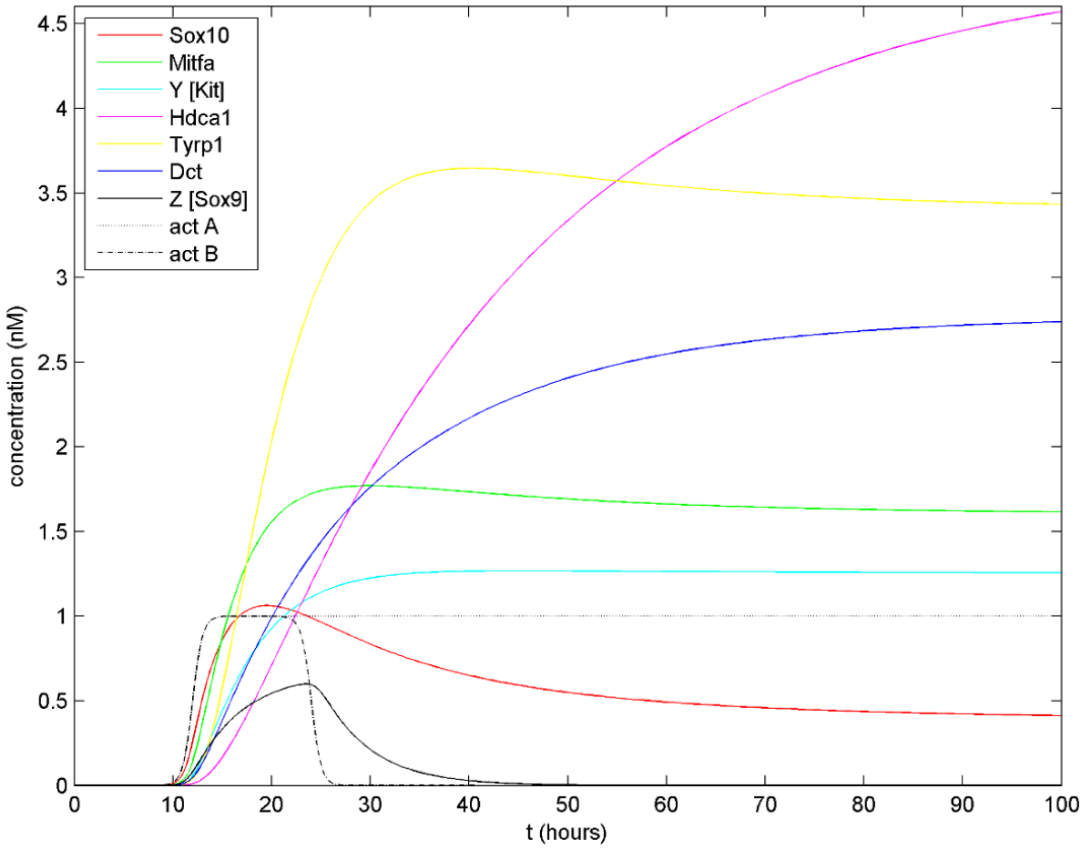

Figure 36. Mathematical model simulation using the Greenhill *et al.* (2011) parameter values.

After the first iteration of parameter optimisation algorithm using the parameter values from Greenhill *et al* (2011), the solutions to the ODEs with the newly found set of parameters were very far from the data points. Therefore some changes were dome in our algorithm to try to get the nonlinear least-squares optimisation algorithm to find parameters that produce ODEs solutions closer to our data points.

First of all we changed the objective error function to a logarithmic one. The new error function is as follows:

$$E = \frac{\log(Y_{meas}) - \log(Y_{pred})}{\frac{\sigma_{meas}}{Y_{meas}}} \qquad (3)$$

In this equation Y_{meas} represents the copies per cell data measured, Y_{pred} represents the copies per cell data predicted by the differential equation solver (ode45) with a set of parameter values, and σ_{meas} represents the standard deviation of the measured data.

We also artificially reduced the standard deviation of the measured data (σ_{meas}) that is used in the objective error function to be minimised. This will force the algorithm to find a set of parameters that produces a predicted data that is more similar to the measured one for the error function to be small. As it can be seen in the error function above (Equation 3), the standard deviation of the measured data is on the denominator of the error quotient. Let's imagine that the data predicted (Y_{pred}) with one set of parameters is far from the measured data (Y_{meas}), which would make the subtraction of the numerator of Equation 3 large. If this subtraction is divided by a large measured standard deviation (σ_{meas}) the resulting error (E) would be small. However, if the subtraction is divided by a small measured standard deviation the resulting error would be large. This would tell to the nonlinear least-squares optimisation algorithm (lsqnonlin) that the current set of parameters produce a solution for the ODEs that is far from the measured data. Hence, the lsqnonlin function will have to keep looking for parameters until it finds a set that minimise the error function. Based on this principle we divided the standard deviation of both the measured *dct* and *mitfa* copies per cell data by a certain factor that would be adjusted in successive rounds of the parameter optimisation algorithm.

A reason why the nonlinear least-squares optimisation algorithm may struggle to find a good set of parameters that produce solutions for the ODE that match the data is that we do not have enough data. The algorithm tries to find optimal values for 50 parameters using just 22 data points. This seems to be insufficient as the algorithm could not find a minimum for the objective error function. To try to overcome this issue we used the smoothed measured data points to construct new data points in between each of the measured data points through interpolation. These new data points will be more densely constructed at early stages (from 20 to 30 hpf) of development. These are the stages where our data changes the most, and where we need a closer match of the data and the predicted ODEs trajectories. We created 22

extra points for each gene, so we had a total of 66 data points to fit, 33 from each gene *mitfa* and *dct*: 11 measured data points and 22 newly constructed data points. Thus, the number of data points was now higher than the number of parameters to fit (66 versus 50), which should help the parameter optimisation algorithm.

After the previous modifications to the algorithm we performed several iterations of the parameter optimisation procedure. As mention before, we used the new parameters found in each round, as the initial parameters for the next one. We also changed the factor by which we modified the measured standard deviation of the data trying to get a better fit of the ODEs trajectories and the data. After several iterations we got a set of parameters (Appendix 9) that produced the ODE trajectories in Figure 37.

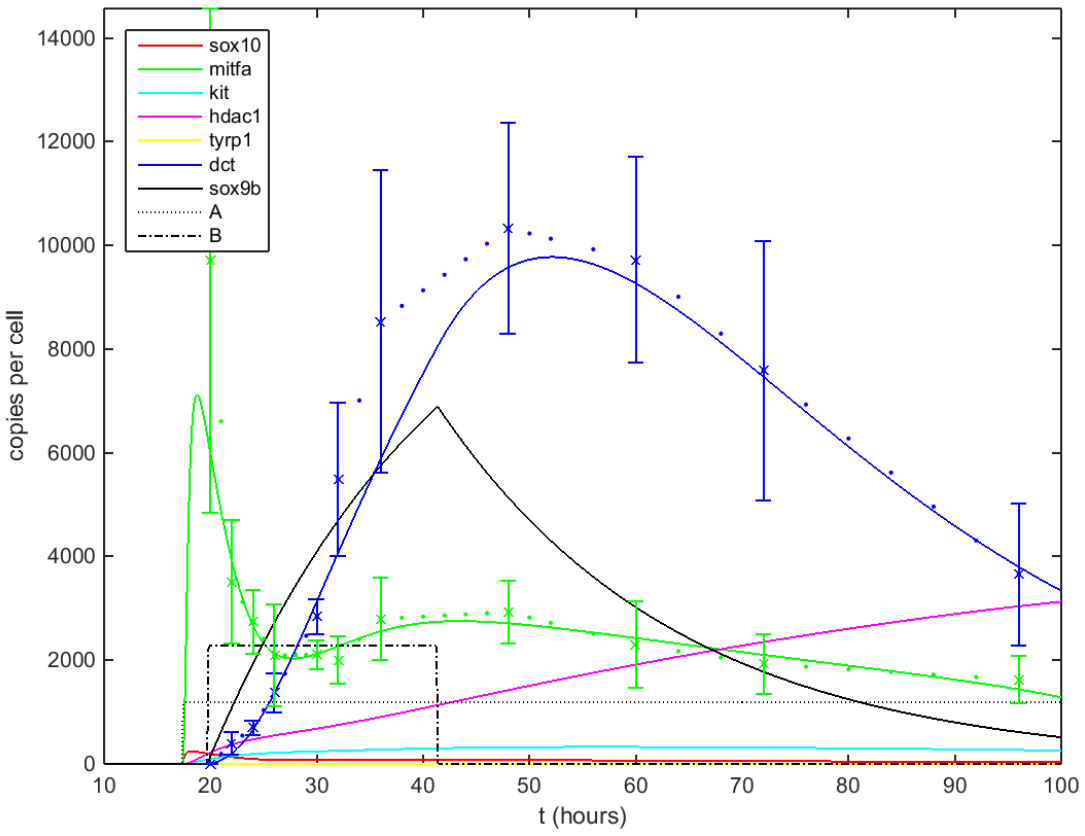

Figure 37. Mathematical model simulation for the new parameters obtained from the parameter optimisation algorithm. The graph plots the solutions for the ordinary differential equations of the model using the new parameter values for each of the GRN genes (solid lines). The figure also shows the measure data points (x) with their associated standard deviation error bars and the constructed data points by interpolation (dots) The new solutions for *mitfa* and *dct* (green and red lines) are very close to the *mitfa* and *dct* experimental data points.

These new parameter values produced the closest trajectories for the *mitfa* and *dct* genes to the experimental data measured. As it can be seen in the figure above, their trajectories fit within the standard deviation error bars for all the data points in both genes. These new parameters have changed the shape of both curves substantially when compared with the trajectories of the Greenhill *et al.* (2011) simulation (Figure 36). Both curves still have their maximum point at approximately the same developmental times: *mitfa* at around 20-24 hpf and *dct* at around 55 hpf. However, in the new simulation *mitfa* expression rises very rapidly early in development, then it goes down substantially and stays down, while in the Greenhill simulation *mitfa* expression was increasing more steadily and stayed up thoughout development. In the new simulation *dct* goes up gradually, like in the Greenhill simulation, but its expression decreases at later stages, unlike in the Greenhill simulation where it stayed up through development.

If we look at the rest of the genes, *sox10* expression is much lower that in the Greenhill simulation throughout. The opposite occurs with sox9b, whose expression is much higher than in the Greenhill simulation. The kick off time for Factors A and B are now different (they started at the same time with the Greenhill parameters). Factor B now stays on for longer than in the Greenhill scenario, and it has became the main factor driving *dct* expression via *sox9b*. The trajectories for both *kit* and *tyrp1* gene expression levels seem oddly low in comparison to the Greenhill simulation. The GRN establishes that Factor A drives *sox10* expression, and then *sox10* regulates *mitfa* activity. However in the simulation above, Factor A, *sox10* and *mitfa* expression seemed to start at same time. When we studied a close-up of the trajectories at early stages of development (Figure 38) it could be seen that Factor A starts expressing first, followed by *sox10* and immediately after, *mitfa*.

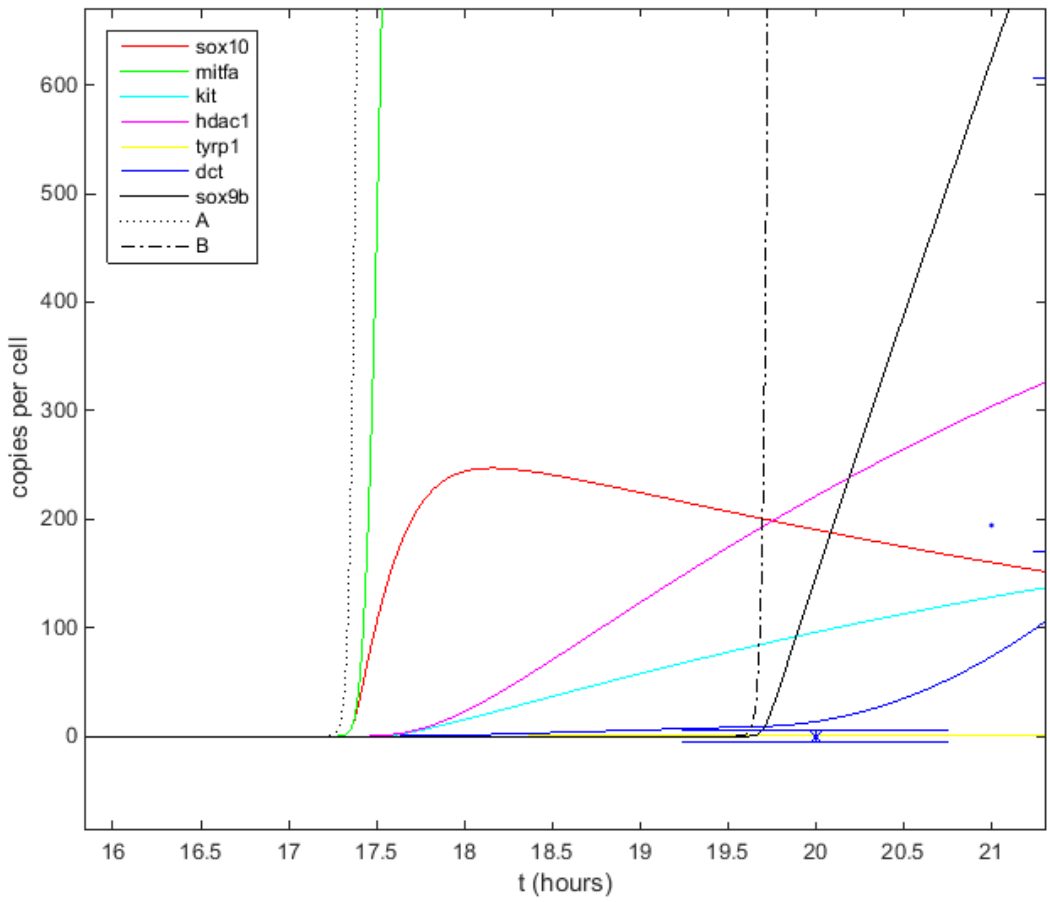

Figure 38. Close-up of the mathematical model trajectories from Figure 37 at early stages of development that show the details of Factor A, *sox10* and *mitfa* induction.

Studying the new parameter values in more detail can help us understand how the model works and why some of the trajectories do not fit the experimental data available. Most of the new parameter values obtained are in the same order of magnitude and are quite similar to the Greenhill et al. (2011) values, but some of them are very different in order to produce trajectories to fit the experimental data. The first new parameter value that is significantly different to the Greenhill et al. (2011) value is β which represent the steepness of the Hill like function that regulates the activity of the input factors A and B. The new value is 50 and the old 2, this is because the new model needs A and B to rise rapidly to activate sox10 and lead to a rapid increase in mitfa expression to reach the high concentrations that we find at 20 hpf in the experimental data. Both β_1 and ϕ_0 have significantly higher values in our new parameter optimisation than the values in the Greenhill model. These parameters control the growth rate of sox10 due to the activation signal from factor

A and Mitfa, and represent the rate of binding and unbinding of the transcription factor to the regulatory element of sox10. The new values of parameters λ_1 and ν_0 are considerably higher than the values from the old model. These parameters are part of the equation that regulates the growth rate of dct due to the signalling coming from mitfa and sox9b. Therefore it makes sense that these parameter are quite high, as the model needs to produce a very high dct expression trajectory to fit the experimental data. The parameter that regulates the growth rate of mitfa expression g_M is also considerably higher in the new model than in the previous one, which is necessary to produce high enough mitfa expression to fit the experimental data. If we consider the parameters that regulate the rate of change of tyrp1 over time, the new value for g_T is significantly lower than the old one, and d_T is a bit higher than before. This would explain why the levels of tryp1 in the new model are so low. The dynamics of hdac1 are similar in both models, as the new parameter values from the equation that describes the dynamics of hdac1 are very similar to the Greenhill ones. The table below (Table 9) shows a detailed comparison of all the parameter values from the Greenhill model and the new values obtained in this project after fitting the model to the experimental data that collected.

Table 9. This table shows a comparison of the parameter values used in the Greenhill *et al.* (2011) model and the new parameter values that we obtained in this project through the parameter optimisation procedure using the experimental data.

Parameter	**Greenhill values**	**New values**
$Z\ (t_0)$ – Sox9b	0	0
$S\ (t_0)$ – Sox10	0	0
$M\ (t_0)$ – Mitfa	0	0
$Y\ (t_0)$ – Kit	0	0
$H\ (t_0)$ – Hdac1	0	0
$T\ (t_0)$ – Tyrp1	0	0
$D\ (t_0)$ – Dct	0	0
β	2	50
t_A	12	17.3836
t_1	12	19.7444
t_2	24	21.5723

A_0	1	0.7879
B_0	1	1.5166
ρ_0	1	1.5163
ρ_1	0.5	0.3125
α_0	2.2	1.7649
α_1	1	1.5014
β_0	1.3	0.5054
β_1	1.2	5.5983
ϕ_0	2	7.5692
ϕ_1	0.1	0.0537
ξ_0	1.6	0.7246
ξ_1	1.1	2.8502
Y^*	0.01	0.1452
M^*	0.01	0.5032
$\gamma_0^{(1)}$	1.5	2.5221
$\gamma_0^{(2)}$	1	3.1304
$\gamma_1^{(1)}$	1	2.5167
$\gamma_1^{(2)}$	0.8	0.8393
k_1	1	3.1280
k_2	1	0.9034
σ_0	1.6	0.5804
σ_1	1.1	2.4771
δ_0	1.6	1.2058
δ_1	0.5	1.7523
θ_0	0.1	0.0290
θ_1	0.5	1.2133
μ_0	0.1	0.0019
μ_1	1.3	1.2893
λ_0	1	0.0337
λ_1	1	51.8319
ν_0	1	15.3247
ν_1	1.2	1.5921
ε_0	0.2	0.1899

ε_1	1.3	1.6547
g_Z	0.1	0.1873
d_Z	0.2	0.0443
g_S	0.3	0.6880
d_S	0.3	3.2927
g_M	0.3	22.8155
d_M	0.3	3.0681
g_Y	0.15	0.0204
d_Y	0.2	0.0998
g_H	0.3	0.3465
d_H	0.03	0.0000
g_T	3.1	0.0165
d_T	0.2	0.3770
g_D	0.3	1.4234
d_D	0.1	0.0736

The new parameter values obtained in the parameter optimisation procedure produce a simulation that predicts high levels of *sox9b* and very low levels of *tyrp1* throughout (Figure 37). However, this is incompatible with the experimental data gathered by Greenhill *et al.* (2011) where expression of *sox9b* was low and restricted just to the early phase of melanocyte development, and *tyrp1* was expressed at high levels at least until 36 hpf (Greenhill *et al.,* 2011).

Therefore we thought that maybe a change in the model equations could produce a better fit to the data. The positive feedback loop between *mitfa* and *kit* should be the main driving force that produce high levels of Dct enzyme, however in the previous simulation is *sox9b* (with a very high expression) the factor that takes up this role, driving enough *dct* expression to make the model fit the data, while *kit* expression is low. Hence we proposed that a change in the *kit* equation, specifically in the way *kit* is activated, could translate in *kit* expression rising to maintain higher levels of *mitfa* and drive high *dct* expression.

In addition, we established on chapter 3 that *sox10* drives *kit* expression earlier in melanocyte development. Hence, we thought that by modifying the *kit* equation to account for this interaction we could achieve enough *kit* expression levels early in melanocyte development to maintain higher levels of *mitfa* and drive enough *dct* expression to fit the data.

Following this reasoning we modified the activation term of the *kit* equation (labelled Y in the model) that describes the variation of *kit* over time, to include the effects of the activation of *kit* by Sox10. The old equation from the Greenhill *et al.* (2011) is equation (4):

$$\frac{dY}{dt} = 2g_Y^M - d_Y Y \qquad (4)$$

The new modified equation to account for the activation of *kit* by *sox10* is equation (5):

$$\frac{dY}{dt} = 2(g_Y^M + g_Y^S) - d_Y Y \qquad (5)$$

In this equation g_Y^M represents the activation rate of *kit* by Mitfa and this term is described in the appendix, as it remains the same as in the Greenhill *et al.* (2011) model. g_Y^S is the new term that describes the activation rate of *kit* by Sox10, and it is described by equation (6) bellow. d_Y represents the degradation rate of Kit. To describe the activation of *kit* by Sox10 we used a Michaelis-Menten type of function and we assumed that the Sox10 transcription factor binds to the *kit* promoter as a monomer (equation 6):

$$g_Y^S = g_Y \frac{\eta_0 S}{\eta_1 + \eta_0 S} \qquad (6)$$

In this equation g_Y represents the rate at which Kit is produced. η_0 and η_1 describe the rates of binding and unbinding of the transcription factor Sox10 to the *kit* promoter. S represents the concentration of Sox10.

In addition, as Greenhill *et al.* (2011) established experimentally that the expression of *sox9b* was low and restricted to the early phase of melanocyte development, and hence the activation of *dct* by Sox9b was close to negligible, we decided to not

include the parameters related to the equation that describes the change of *sox9b* expression over time in the model. In this way we may avoid the parameter optimisation algorithm to use *sox9b* as the main driver of *dct* expression, and it will be forced to look at the other mayor diver of *dct* expression, *mitfa*, when finding new parameter values.

It was also noticed that the 20hpf data point for *mitfa* had very large standard deviation bars and that it may have been overestimated. This might have been because *mitfa* is expressed in many cells but under the detection levels of the *in situ* hybridisation techniques at 20 hpf. It might be that the expression per embryo trunk was very high but the number of cells expressing *mitfa* that were detected at that time point was low, as many *mitfa*-expressing cells would be under the detection threshold and would have not been counted. Both of these pieces of data produced a copy number per cell estimation that could be aberrant. As a consequence, it seems reasonable to penalise less those parameter value combinations that produce a *mitfa* trajectory that is not that close to the *mitfa* data point at 20 hpf. To do that its standard deviation was artificially made 100 times bigger than the standard deviation of the rest of the data points. Hence, those models that produce a curve for *mitfa* which does not come close to the 20 hpf data point will not be too penalised, as this point might be an outlier.

Thus, we performed these changes in the parameter optimisation algorithm and ran the procedure again in Matlab to see if a better parameter values is achieved, and hence better trajectories in the model. The optimisation algorithm was run for 17 iterations and the trajectories obtained were the ones in Figure 39.

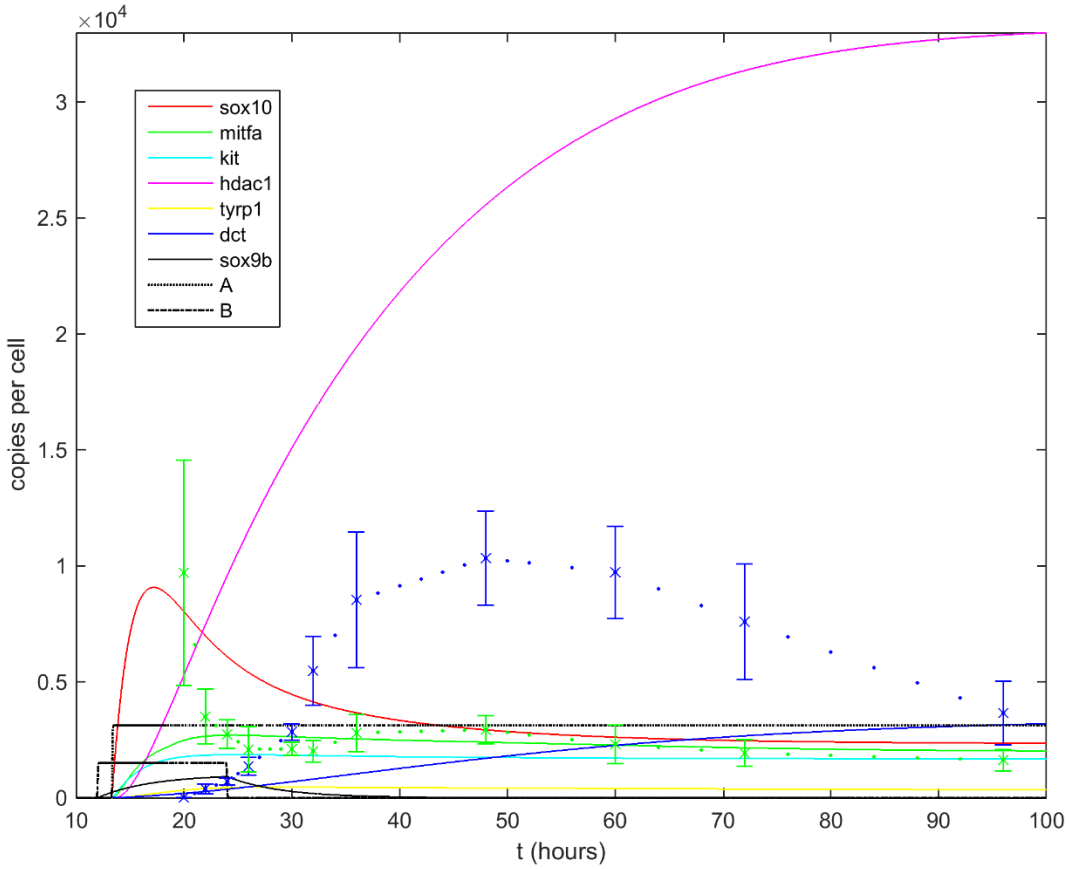

Figure 39. Mathematical model simulation for the new parameters obtained from the parameter optimisation algorithm after making modifications in the *kit* equation, *mitfa* 20 hpf data point, and *sox9b* parameters. The graph plots the solutions for the ordinary differential equations of the model using the new parameter values for each of the GRN genes (solid lines). The figure also shows the measure data points (x) with their associated standard deviation error bars and the constructed data points by interpolation (dots).

After making these changes, the trajectory for *sox10* is higher than before and it is closer in shape to the one in the Greenhill model. *mitfa* now does not go up sharply to try to reach the data point at 20 hpf, as we have modified its standard deviation, and its trajectory is quite close to the *mitfa* data points if we do not take into account the 20 hpf point. *kit* expression seems to be higher than previously. *tyrp1* expression is still very low, and due to the changes made we have decreased the expression levels of *sox9b*, although it might have been too much, as its trajectory is much lower in relative terms than that in the Greenhill model. Unfortunately, these changes have not achieve a *dct* trajectory that goes close to the data points. *hdac1* trajectory shoots off very much like in the Greenhill model.

Discussion

Our aim was to develop a quantitative mathematical model to describe the gene regulatory network that drives melanocyte development. To produce a quantitative mathematical model we need to collect data so we can generate the equations and parameter values to describe that data. Therefore, if we want to model a single cell behaviour we should measure single cell data. However, measuring single cell gene expression data is very challenging. Our initial idea was to use a transgenic line that uses the *sox10* promoter to drive EGFP expression with a FACS machine to select *sox10* expressing cells, that then we would use to measure gene expression. However a *sox10*:EGFP transgenic line would label many other cells besides melanocytes, so a better line to use would be *mitfa*:EGFP, as *mitfa* only labels cells with potential to become melanocytes. With this approach we would count the number of melanocytes sorted and then we would use them to extract RNA and measure the gene expression of the GRN players through qPCR. We would get a measurement of gene copy number per cell. Unfortunately at the start of this project the laboratory did not have local access to a FACS machine able to do sorting.

Therefore we came up with a different experimental set up. We would obtain an estimate of gene expression per melanocyte measuring the gene expression of the 4 genes we wanted to analyse (*sox10*, *mitfa*, *kit* and *dct*) in 5 embryo trunks, and then we would do four ISH for the same genes in sibling embryos to count the number of cells expressing that particular gene in the trunk. In this way we would be able to produce an estimate of gene copy number per melanocyte that could be used to perform the parameter optimisation.

Copy number quantification

Three of the primers that we used to amplify our four gene of interest (*sox10*, *kit* and *dct*) were designed by the company PrimerDesign and described in Table 1 in Chapter 2. We chose this company to design the primers because we got a sponsorship from them, but also because they guarantee primer specificity and an efficiency of more

than 90%, as well as validating the primer on zebrafish embryo cDNA that I sent them. Therefore we were sure the primers we got were of very good quality and were going to work in our samples. They did not design the *mitfa* primer for us because we used a primer previously designed by Laura Vibert in her thesis.

The qPCR reaction produced no amplification in any of the NTC (with water and reaction mix, instead of template and reaction mix) wells for all the four genes, which means that there are no primer dimers or contamination that would lead to non-specific amplification. The melt curves of all the genes showed a single sharp peak, which means that only one amplicon was amplified, and no primer dimers or genomic DNA amplification was produced, so our primers are very specific for our amplicons.

Every stage was measured in triplicate for each gene, and a mean and standard deviation of the three measurements was produced. Then the qPCR program (Illumina Eco software) used the standard curve to produce a copy number average and its associated standard deviation. This average value was measured from 5 embryo trunks, so we divided it by 5 to calculate the copy number per trunk for each stage and gene, and this is what we plotted in Figure 23 to Figure 26.

These data showed that there are some big jumps on gene expression in most of the genes between contiguous stages. This happens in *mitfa* (Figure 25) for example, where gene expression is high at 18 hpf then it goes down drastically at 20 hpf, to then come back up again at 22 hpf. This is unlikely to happen in reality, especially at the beginning of development when we sampled every 2 hours. This most likely represents noise in the gene expression measurement that could be caused by several causes: embryo-to-embryo gene expression level variation, marginally different staging of the embryos within the same time point, slightly different dissected trunks used for the measurements, or variation in the amount of cDNA incorporated into each of the qPCR triplicates. That is why we decided to smooth the data to reduce the influence of noise (Spiess *et al.*, 2015). By smoothing the data points of a signal are modified taking into account the signal of the contiguous data points. So the points with a signal that is higher than that of the adjacent points are reduced, and points that have a lower signal than the nearby ones are increased. This reduces the noise of the signal making the data smoother. The method that we

selected to perform the data smoothing in Matlab was the Savitzky-Golay, which has been used before with qPCR data (Spiess *et al.*, 2015).

After the smoothing procedure the noise has been reduced and the data seems to show less jumps in gene copy number in between contiguous stages. However, there are still some dips in the copy number of some genes such as *mitfa* at early stages that do not look very likely in a biological context. The cell counts would shed some light into the copy number results and will help us make more sense of it.

Cell number count

We had measured the gene expression levels for each gene in five embryo trunks and have estimated the gene copy numbers per stage in one trunk. In order to calculate the gene copy number per cell on the trunk we need to know the number of cells expressing each of the genes that there are in an embryo trunk. For that we carried out ISH for each gene and stage.

In the lab we normally use NBT/BCIP to label the signal in ISH, but in this case we used FastRed, which emits red fluorescence under a fluorescent microscope. We decided to use FastRed because fluorescence signal is much easier to identify and distinguish from the background by software. This is because a fluorescence image is captured in black and white where intense white pixels represent signal, black pixels indicate lack of signal, and background is denoted by pixels with low intensity values of white. This is easier to identify by software than a colour image from a NBT/BCIP ISH where purple would represent signal but at the same time there are many other confounding colours on the image representing background and other structures in the embryo. Also FastRed has the advantage that it works very similarly to NBT/BCIP: the specimens have to be incubated with the FastRed solution, which the alkaline phosphatase will transform and deposit in the gene expressing cells. This means that the whole ISH protocol is just the same as the conventional one, just changing the staining solution.

Initially we thought about using the High Content Microscope (HCM) to image all our specimens. We have four genes, with 12 stages each, and we wanted to count four

ISH embryos per stage, which make a total of 192 embryos to image and count. The high throughput capabilities of the HCM made it ideal for our purpose. The initial idea was to put each of our embryos in one of the wells of a 96-wells plate and then image them using a Z-stack to capture the whole width of the trunk. However, when I did a test with just some embryos to learn how to use the microscope we realised that it was not going to be as simple as that. If we randomly placed our embryos in glycerol to the well it was very likely that they were not position laterally but even more important it was very unlikely that they were right at the bottom of the well, so each embryo would be in slightly different plane relative to the bottom of the well. The HCM has a setting to account for that Z offset (if the sample is in a different plane than the bottom of the well), however once you set this offset it is the same for the 96 wells of the plate, and in our case each embryo would be different. Therefore we came to the conclusion that the only way to make this work was to mount each embryo individually in a lateral position and at the bottom of the well. For that we could not use 96-well plates, as there was not enough room to mount the embryos so we used 12-well plates. Nevertheless, even though we placed the embryo right at the bottom of the well, there was enough difference in between the embryo in one well to the one in the next one so that we could not use the same offset parameter. Thus we realised that it was not going to be possible to automate the image acquisition in a time effective way and that we needed to do it one by one. This meant that we could not use the major advantage of the HCM (high throughput and automation), and if we were going to use it like a normal microscope it was difficult to operate as all the controls were through the software. Moreover, even when we managed to take a photo of our embryos, the background was quite high as the system could not eliminate the out of focus signal from the rest of the planes. This made the images very difficult to analyse using the microscope software, and we were never capable of successfully segmenting the melanocytes in the embryo trunk to count them. Finally, we concluded that the best option was to use a stereoscope to acquire high quality images of our embryos and then analyse those images with a software to count the melanocytes.

At this time, we acquired a Zeiss Axio Imager 2 fluorescence microscope with the Apotome technology in the lab. The Apotome technology allows you to perform optical sectioning in fluorescence images, preventing the scattered light from out of focus planes in thick specimens. It produces images with high contrast and resolution and very low background. The main advantage is that is as easy to use as a conventional microscope. We decided to use this technology to take our images and then analyse them with a software called MINS.

To take the images we mounted the specimen on a slide and coverslip and used the Zeiss Axio Imager 2 microscope with the Apotome module to take Z-stack images. We used the 20x magnification objective, the one with the lowest numerical aperture that fitted the whole trunk region in a single plane, to focus on the region of the trunk we needed to count (between the beginning and end of the yolk sac extension), and imaged the whole thickness of the embryo using Z-stacks.

We initially intended to analyse these images with the software MINS (Modular Interactive Nuclei Segmentation), which had been previously used to analyse and count stem cell cultures and early stage mouse embryos (Lou *et al.*, 2014). The program is designed to segment nuclei so when we tried to use it to detect ISH labelled melanocytes in our embryos the program failed, as the melanocytes morphology is more complex than the nuclei one. Therefore we thought that maybe by staining the nuclei of all the cells of the embryo with Hoechst staining the program would be able to detect all the nuclei in the blue channel and then expand the nuclei area and look for signal in the red channel. Then it would just count the cells with signal in the red channel, which are the melanocytes.

We tested this first with one embryo, and to get enough resolution for the program to segment the nuclei, we had to image the embryo with the 63x objective. The nuclei segmentation did not work when using the whole Z-stack as the images were too heavy. We reduced the number of nuclei present in the Z-stack by using a subset of 4 planes from the original Z-stack and we tried to analyse it again. It took almost 48 hours, but we got a result with all our nuclei segmented quite nicely (Figure 29 B). However when we tried to expand the nuclear area and look for signal in the red channel the program detected way more melanocytes that there actually were on

the image, as it assumed that some of the background (even as low as it was) was real signal coming from the melanocytes. Moreover, this segmentation took several hours again. That is why we concluded that the most time efficient and straightforward way to count our cells was to do it manually by eye.

In order to count manually the cells we used a plugin from ImageJ that allows you to click a cell and put a marker on it so you can see the ones you have already counted in an image plane. This marker helps you see if a cell in the next plane is the same one that you have already counted in the previous plane, as the marker will be in the same position. Counting the cells in this way was labour intensive and you needed to be focused to avoid counting the same cell twice. Because there is a lot of room for human error while counting, we performed the counts randomly i.e. we selected the stages to count randomly so we reduced the human expectation of counting more cells in older stages. Also we counted every image three times in the case of *dct*, and twice in the case of *mitfa*, *sox10* and *kit* at different points in time. By doing this we try to reduce the human error in the counts as much as possible.

We started counting *dct* cells, as this gene is a melanocyte differentiation marker and provides the best estimate of melanocytes at each stage (Elworthy *et al.*, 2003; Greenhill *et al.*, 2011). We imaged four specimens per stage and we counted three times each image. As it can be seen in Figure 31, we could not count any *dct* expressing cell in the trunk at 18 hpf, and we could detect very few at 20 and 22 hpf. However the number of *dct* expressing cells was high by 26 hpf.

In order to evaluate how accurate our *dct* cell counts are, and how well they represent the melanocyte number we decided to count melanised melanocytes in fixed WT fish at 60, 72 and 96 hpf stages. The number of melanised cells at 60 hpf is slightly higher than the number of *dct* expressing cells counted (71 vs 60 average). The number at 72 hpf is very similar (58 vs 56). And the number at 96 hpf is considerably higher in melanised melanocytes (89 vs 54). Even though some of the numbers are higher in melanised melanocytes they are in the same order of magnitude, and some are very similar. That suggest that the number of *dct* expressing cells represents a good estimate of the real number of melanocytes. We may be under-counting *dct* expressing cells because when we counted melanised

cells the embryos were treated with epinephrine prior to fixation, which contracts the melanosomes into the cell body, making the cells smaller and easier to distinguish (Johnson *et al.*, 1995). This is likely to explain why the number of melanised cells is slightly higher in some cases than the number of *dct* expressing cells.

Mitfa is a better melanocyte marker for early stages of melanocyte development (20 or 22 hpf) in the embryo trunk (Montero-Balaguer *et al.*, 2006; Nguyen *et al.*, 2010). Therefore we started counting *mitfa* expressing cells until the cell counts match the *dct* cell count. From that point onwards (32 hpf) we assume that *dct* is a better melanocyte marker (Figure 32).

Johnson *et al.* (1995) carried out some melanocyte counts in whole fish body, where they count all the melanocytes. Their first data point is at half a week after fertilisation, and they counted around 400-500 cells (exact number was not known as data was only provided in graphic form). This stage would be similar to our 96 hpf data point, where we counted around 60 *dct* positive cells or 89 melanised melanocytes. Taking into account that we only counted a section of the trunk which might represent approximately a third of the whole embryo, our measurements seem to be slightly lower, but still within the same order of magnitude to theirs, which is reassuring (Johnson *et al.*, 1995).

Sox10 is expressed in many other tissues beside melanocytes. Therefore just a part of the *sox10* expressing cells counted would be melanocytes. Also the *sox10* expressing cells are very close together and they are not well defined, which make very difficult to distinguish one cell from another while counting. We counted *sox10* expressing cells in half of the stages (six) but we realised that we would not be able to use these numbers, as they do not represent only melanocytes, which are the cells we are focusing in our GRN.

The same happened with *kit*. It is expressed in many other tissues other than melanocytes. Only some of the *kit* expressing cells will be melanocytes. The main confounding cell type in this case were the mucus secreting cells. These are small and round cells located in the epidermis of the whole fish in a very high number. When we counted the cells laterally was relatively easy to avoid counting the first and last

two or three planes at both sides of the embryo which contained the epidermis and the mucus secreting cells. However, these cells were also present dorsally in every plane throughout the embryo, and it was very difficult to distinguish them from melanocytes when counting. We counted half of the stages (six again) but came to the conclusion that we could not use these counts to estimate melanocyte numbers.

As we concluded before, we could not use the *sox10* and *kit* qPCR copy number data for the same reasons: melanocytes would contribute to those numbers only in part, with the rest being expressed by many other cell types. Therefore we used just the *mitfa* and *dct* copy number data for the copy number per cell estimation. The cell number count needed came from the *mitfa* expressing cells counts from 18 hpf to 32 hpf, and the *dct* expressing cell counts from 36 hpf onwards.

As I did with the copy number data we performed a smoothing procedure to reduce the noise in the data that may have been caused by errors in the counting. We used the same algorithm as in the copy number smoothing and the resulting data looked very similar to the original data although slightly smoother.

Copy number per cell estimation

The GRN mathematical model describes the behaviour of a single melanocyte. Hence, to perform the parameter optimisation we need the copy number per cell data for every stage of both genes. To get this we need to get the quotient of the copy number per stage and the number of cells we counted at that stage for both genes.

To get the quotient of the standard deviations (ΔC) we need to use the following formula:

$$\Delta C = \frac{\bar{Y}\Delta X + \bar{X}\Delta Y}{\bar{Y}^2}$$

Where $\bar{Y}$ represents the cell number, ΔX is the copy number standard deviation, $\bar{X}$ is the copy number and ΔY is the cell number standard deviation.

Figure 35 shows the copies per cell of both *mitfa* and *dct* throughout the different stages measured. *mitfa* seems to have a high copy number per cell at early stages of development (20 hpf) and the copies decrease quite rapidly and plateau. Dct copies per cell start low at early stages of development, then grow to reach a maximum at around 48 hpf and then come down towards the later stages of development measured. The result for *mitfa* at 18 hpf was unusual. When we produced the quotient of copy number and cell number, our Matlab script assigned a value of zero copies per cell for the 18 hpf time point because the result was undefined. That is because at 18 hpf we have *mitfa* expressed at a fairly high copy number but we could not identify any *mitfa* expressing cell in the trunk (Figure 31). This might be because there is widespread *mitfa* expression that is below the detection limit of the ISH, or maybe it was contamination. A repeat of this experiment would certainly help to determine if this is an artefact or is real widespread low expression. For *dct* both copy number and cell number were 0 at 18 hpf. Therefore, we decided not to use the 18 hpf time point of either of the genes for the parameter optimisation, as it is probably erroneous.

It seems odd that the number of copies per cell of *mitfa* at 20 hpf is so high, and then it decreases by approximately 6000 copies per cell in just two hours. We would expect changes taking place in a less abrupt manner in biology. However this data point (20 hpf) has a huge standard deviation, so it may be that if we repeat this measurement, the new number of copies per cells is in the bottom part of the standard deviation (around 5000 copies per cell) and then the change to the next stage would be smoother and make more sense. The number of copies per cell is fairly stable for the rest of the development at around 3000 copies per cell. This agrees with the data suggesting that *mitfa* expression stays on in melanocytes though development, as there are some structural genes for melanosome biogenesis or melanin synthesis that are driven by *mitfa* after melanocyte specification (Johnson *et al.* 2011).

It has been shown that *dct* is expressed immediately after *mitfa* in melanocyte development (Opdecamp *et al.*, 1997b; Jiao *et al.*, 2004), so we would expect to see *dct* expression rising fast at around 24 or 26 hpf, just after *mitfa* expression peak

(Figure 35). However, this does not seem to be the case if we look at our *dct* copies per cell plot: *dct* expression starts rising fast at around 30 hpf, reaching maximum expression at around 48 hpf, which is slightly later than expected. Having said that, if we look carefully at the scale of the plot we can see that at 24 or 26 hpf we have around 1000 or 1500 copies per cell, which is quite significant, and very similar to the number of *mitfa* copies per cell if we omit the 20 hpf time point. Then *dct* expression seems to shoot off and reach very high numbers at 36, 48 or 60 hpf. However, it is worth pointing out that the standard deviation of these time points is fairly high, so it may be that when repeating the experiment the number at these stages come down to a more lower level.

The copy number per cell data is not a direct measure of gene expression in a single cell. It is an estimate obtained from a copy number measured by qPCR and a cell number count from ISH images. The copy number data measured by qPCR is also an estimate as initially we measured the gene copy numbers in five embryo trunks and then assuming they have the same gene copies we divided by five to calculate the copy number per trunk. At the same time the five trunks per stage that we used to extract RNA and measure gene expression were dissected manually, so it is quite likely that there are some trunks slightly bigger, which means they will have more copies of our gene of interest. Hence there are lots of opportunities for error and to introduce noise in this whole process that will make our final gene copy number measurement per trunk not very accurate. Ideally, we would have measured the gene copy number in cDNA obtained from a single embryo trunk, but we did not manage to extract enough RNA material from one trunk to measure it through qPCR.

We had a measurement of gene copy number in a trunk, so we needed to know the number of cells expressing that gene in the trunk. We could not use the same trunk to perform the qPCR and to count the number of cells. Therefore we used sibling embryos, fixed at the same time as we extracted the RNA from the other trunks, and assumed they will have the same cell numbers. It is quite unlikely that one embryo has exactly the same number of melanocytes than another, so we decided that in the same way that we were measuring by qPCR 5 embryos and then divide by 5 to get the average gene copy number in one trunk, we would do the same for the cell

counts. We counted 4 embryo trunks and then we obtained the average number of gene expressing cells for one trunk. We counted the embryos from a microscope Z-stack image, so it was likely that the trunk section that we counted was slightly different to the section we used to measure gene expression by qPCR. At the same time, when counting Z-stacks there will be cells that extend over more than one plane, so we had to keep track of the cells that we counted in one plane and avoid counting them in the next one. This would introduce some error in our counts, to try to avoid this we counted every image two or three times and we had four embryos per stage; so we averaged the number of cells from all the four embryos and the repeated counts to obtain the number of gene expressing cells per stage.

Therefore, in the process of obtaining the copy number per cell data that we needed to perform the parameter optimisation procedure we made a lot of assumptions and introduce quite a lot of errors. But nevertheless, it is likely to be the most accurate, thorough and more densely populated measurement of gene expression of zebrafish melanocytes available to date, which we think is a big step forward.

Ideally we would have liked to measure directly gene expression in single cells. For that we could use Percoll density gradient and a subsequent FACS procedure to sort out melanised cells, as they will be the only ones absorbing light when excited by the laser following a similar protocol as the one used by Higdon *et al.* (2013). However, if this method is used melanocytes could only be isolated from around 26-28 hpf when they become melanised (Karlsson *et al.*, 2001), meaning that we would not be able to isolate melanocytes prior to this stage. To overcome this issue, we could set up a transgenic fish with *mitfa* promotor driving EGFP expression which would label melanocytes from early stages. Then FACS could be used to sort green fluorescence single cells, whose gene expression measurements could be done using the Nanostring nCounter technology. This technology is capable of detecting and count a specific nucleic acid in a complex mixture. A multiplexed probe library is made with two sequence-specific probes for each gene of interest. The first probe, a capture probe, contains a sequence complementary to a particular target mRNA plus a short common sequence coupled to an affinity tag such as biotin. The second probe, the reporter probe, contains a sequence complementary to the target mRNA, which is

coupled to a color-coded tag that provides the detection signal. These differently coloured probes create a unique code for each gene of interest and allows the detection of hundreds of targets at the same time. Each target molecule of interest is identified by the colour code present on the reporter probe. The level of expression is measured by counting the number of codes for each mRNA. This technology offers high levels of precision and sensitivity being able to detect even less than a copy of the gene of interest per cell (Geiss *et al.*, 2008). However this technology is very expensive, and the budget or equipment was not available to do it in this project. Still single cell genomics is the direction that our lab is taking.

Parameter optimisation

Greenhill *et al.* (2011) produced a mathematical model together with a set of parameters that described in a semi-quantitative manner the development of a single melanocyte. When solving the ordinary differential equations using the parameters they obtained the plot in Figure 36 is produced. This graph represents how the gene expression of the different players of the network varies along development. In order to select the parameters for their model they used ISH expression data and information on the expression of the GRN genes from the literature to find a set of parameters that produced the curves with a shape that matched the data available. Hence the concentration and exact measurements predicted by the ODEs are not very accurate.

Our aim was to begin to make this model quantitative, and for that we needed to modify the parameters so the ODEs predicted trajectories match the data we collected as accurately as possible. The first issue we faced is that the units of our data were different to the units that Greenhill *et al* (2011) used in their model. The data that we collected is measured in gene copy number per cell, whereas their data is measured in μM units. We think that our units are more intuitive to represent absolute gene expression because when using molar units one needs to know the volume of a melanocyte, which is fairly difficult to measure. Therefore the units of the parameters selected by Greenhill *et al* (2011) were in the order of one, and if we

try to use our parameter optimisation algorithm to modify those parameters to fit our data it will never work, as the units of our data are thousands of times higher than the units of the parameters. So we had two options, to modify the Greenhill parameters so they are in similar units to our data or to scale down our data to have it in similar units to the Greenhill *et al.* (2011) parameters. The rigorous way to do it would be to scale the Greenhill parameters, but this cannot be done using the same scaling factor for all of them, as some parameters represent initial gene concentrations, some are gene kick-off times, some are degradation rates, and other TF binding rates. Hence the scaling for each parameter would need to be done individually, evaluating the equation in which the parameter is acting. We decided that the second option was more straightforward, as Greenhill *et al.* (2011) already probed that those parameters produce a set of trajectories for the mathematical model that matched their data.

When the first round of the parameter optimisation procedure was done the trajectories of *mitfa* and *dct* were very far from the data. One of the main reasons for that is that the nonlinear least-squares optimisation algorithm was trying to find values for the 50 parameters of the model that minimised the objective error function by using only 22 data points. We did not have more measured data, but we used interpolation to construct new data points in between our measured data points. We produced data every hour from 20 to 30 hpf, every two hours from 32 to 52 hpf and every four hours from 56 to 96 hpf. Our newly constructed data points were denser at early stages of development because our measured data was also denser at this stages. We measured the gene expression levels every two hours until 30 hpf, which meant that we only needed to create a new data point in between our measured data points by interpolation for these stages. Also, the gene expression levels of our measured genes changed more at early stages of development, so we needed more data points at these stages if we wanted an accurate match of the predicted trajectories to the data. We created 22 data points extra for each gene, making a total of 66 data points for both *mitfa* and *dct*. With this new data the optimisation algorithm had to fit 50 parameters using 66 data point, which produced better results.

Another way to force the optimisation algorithm to find parameters that produce trajectories closer to the *mitfa* and *dct* data was to reduce the standard deviation measured for both genes. This standard deviation is part of the objective error function that the lsqnonlin algorithm needs to minimise. Hence, a smaller measured standard deviation will penalise more the predicted trajectories that are very far from the measured data.

These modifications proved successful as the parameter optimisation algorithm found a set of parameters that produced trajectories for *mitfa* and *dct* genes that were very close to the data after several iterations of the algorithm. *sox10* expression levels are much lower in the new simulation than in the Greenhill one. This was unexpected as *sox10* is the driving factor of *mitfa*, whose expression was high. However, our model suggests that low quantities of *sox10* are sufficient to drive enough *mitfa* expression to fit our data. Moreover, the low level of *kit* expression they seemed to be enough to maintain enough *mitfa* expression through the feedback loop to fit the data.

The simulation suggests that Factor A, *sox10* and *mitfa* expression start at same time, although our GRN states that Factor A activates *sox10*, which then drives *mitfa* expression. When we studied a close-up of the simulation trajectories at early stages of development, we saw that indeed Factor A started expressing first, followed by *sox10* and immediately after, *mitfa*. We expected a delay in between expression of *sox10* and activation of *mitfa*, however this model assumes that the binding of TF to DNA to regulate protein production is a single step. And the ODE equations that describe the gene expression changes are built in a way where they only need a few molecules to start the exponential growth. That is why we see a huge increase of *mitfa* with very few *sox10* molecules.

The main feature that our model with new parameters highlighted is that for *dct* to be expressed at such a high values, the *sox9b* activation of *dct* needed to be much higher than in the Greenhill simulation. This was achieved by increasing the expression levels and expression timespan of factor B, which drives *sox9b*. Hence, contrary to the Greenhill *et al.* (2011) model where the effect of *sox9b* activation of *dct* was restricted to the early phase of melanocyte development, our model

suggests that this interaction last for longer and become a very important activation route of *dct*. This is a very interesting outcome of the model that makes specific predictions that need to be experimentally tested. For that we need to obtain gene expression levels of *sox9b* throughout development. If the measured *sox9b* expression levels match the *sox9b* trajectory predicted by of our model, it means that *sox9b* plays a more important role in melanocyte differentiation than Greenhill *et al.* suggested. If however, the experimental data confirms that *sox9b* expression levels are much lower than the ones predicted by the model, it means that our model is wrong and hence the parameter values obtained would need to be changed again. We would need to use those *sox9b* measurements in our parameter optimisation algorithm to find new parameter values. The new resulting model will need to find a new interaction that can explain the high *dct* expression levels measured in the data.

Our model also predicts very low levels of tyrp1 throughout. This was unexpected as ISH data from Greenhill *et al.* (2011) demonstrated that *tyrp1* is expressed at high levels at 36 hpf. This would mean that our model needs to be improved further, and the measurement of *tyrp1* expression levels throughout development would be a key dataset to perform the next rounds of parameter optimisation.

Both *sox9b* and *tyrp1* gene expression measurements would optimise the parameter set to obtain a mathematical model that explains the data better. This would very likely highlight new genetic interactions to explain the data that would involve the modification of the GRN, or it would confirm that that the current GRN is a good model for melanocyte development.

A detailed look at the new parameter values obtained from the parameter optimisation procedure can help us understand some of the new features of the model. The parameter β that controls the steepness of the function that regulates the activity of the input factors A and B is too high, making the function gradient very sharp. This is very unlikely to happen in nature, where the gene expression grows more smoothly. However this sharp step was necessary to activate very rapidly the expression of *sox10*, so that *mitfa* expression could increase enough to be close to the first data point at 20 hpf. As we need *mitfa* expression to go up fast and quite high some of the parameters regulating its expression growth rate have increased significantly. The same happened with the

parameters involved in the equations that describe the dynamics of *dct* expression. Both *mitfa* and *dct* expression has to be very high to fit the experimental data collected, and hence the values of the new parameters are considerably higher. The new values are between 20 and 50 times higher than in the Greenhill *et al.* (2011) model for λ_1, ν_0 and g_M, which is probably unrealistic as it is quite unlikely that some of the rates of transcription factor binding and unbinding to the gene promoter in the same model change from approximately 1 to around 20 or 50. The value of most of the new parameters remain in the same order of magnitude to the ones in the Greenhill model, however some parameters had to be changed quite significantly by the parameter optimisation algorithm to fit the trajectories to the data collected, whose expression levels were higher than those displayed in the Greenhill model.

Experimental evidence from Greenhill *et al.* (2011) showed that *sox9b* is expressed at low levels and just at early stages of melanocyte development, and *tyrp1* is expressed at high levels throughout. Hence we attempted to change the model to try to achieve a better fit to the data. The equation that regulates the change in *kit* expression over time was modified to include a term that describes the activation of *kit* by Sox10, as it was experimentally tested in chapter 3. This term is described by a Michaelis-Menten type of function that describes the activation of *kit* by Sox10 taking into account the binding and unbinding rates of the transcription factor to the *kit* promoter. This type of function is quite widely used to describe binding of transcription factors to regulatory elements (Greenhill *et al.*, 2011; Leal Valentim *et al.*, 2015; Aguilar-Hidalgo *et al.*, 2013).

The *sox9b* parameters were excluded from the parameter optimisation to avoid that they are used as the main driver of dct expression because this is not what happens in the experiments. The first experimental data point of *mitfa* expression seems to be aberrant, and hence its weight in the algorithm was modified so trajectories that do not come close to it are not penalised very much.

This changes made *mitfa* to come closer to the experimental data collected, if the 20 hpf data point is not taken into account. However the fit of the *dct* trajectory to the experimental data is very poor. The trajectory does not come even close, only being close in the last time point at 96 hpf. The *sox10* trajectory seems to be better after the modifications, as it goes up rapidly early on, driving *mitfa* expression and *kit*. *kit*

expression is higher than in the first model, which is probably due to the new term in the equation that accounts for the Sox10 activation of *kit*. *tyrp1* expression is still low, but it should be quite high according to the Greenhill *et al.* (2011) experimental data. *hdac1* shoots off like in the Greenhill model. The modifications have also reduced the expression levels of *sox9b*. It may be that *dct* expression is that low, because we have not allowed *sox9b* expression to rise to drive *dct*.

This highlights that the new model with the changes is not much better than the previous one. It certainly produce better trajectories for some genes such as *sox10*, *mitfa* or *kit*. But some other trajectories like *dct* or *tyrp1* are still very far from the experimental data. It might be that going through more iterations of the parameter optimisation procedure could help to refine the parameters and bring the gene trajectories closer to the data. However the trajectories did not change very much in the last 4 iterations of the parameter optimisation algorithm that were carried out. Hence, it may be that the algorithm has found a local minima and is looking for parameter values in the wrong place of the parameter landscape, which is the n-dimensional space made up of all the possible parameter value combinations. To try to force the algorithm to look in a different place of the landscape we could randomly change the values of the parameters and run the algorithm in an iterative way to check if it produces better trajectories. The other alternative would be to use Monte Carlo simulation, which is a method to sample randomly the parameter space by varying the parameters and checking if the trajectories are close to the data points (Matlab documentation). This is a very powerful method as it would search the whole parameter landscape, and it is less likely to get stuck in a local minima. However due to time constrains we were unable to test it in our project. This, together with obtaining data from more genes and more time points, would certainly help to find a parameter value combination that can produce trajectories that fit the experimental data, making the model more powerful and capable of producing predictions.

Conclusions

Greenhill *et al.* (2011) developed a core GRN underlying melanocyte specification and differentiation in zebrafish. They used semi-quantitative ISH data to build a mathematical model that explains melanocyte development and makes predictions about the gene expression changes during melanocyte differentiation. We sought to improve this mathematical model making it quantitative. For that we measured gene expression data throughout development to produce the parameter values that will allow the ordinary differential equations of the model to describe that data.

We measured the gene expression levels of *sox10*, *mitfa*, *kit* and *dct* at 12 stages along development in 5 embryo trunks through qPCR. We also did four ISH for those genes with sibling embryos for the 12 stages and counted the number of cells expressing each gene in the embryo trunk. We used both measurements to produce an estimate of gene copy number per melanocyte that was used to perform the parameter optimisation of the mathematical model.

The parameter optimisation aims to find the best parameter values for the mathematical model so its predictions are as close as possible to the experimental data collected. For that we developed a Matlab algorithm that was based on the use of an objective function that calculates the error from the predicted trajectory calculated with a set of parameters and the measured data; and we try to minimise that error finding new parameter values through a nonlinear least-squares solver.

The parameter optimisation algorithm initially struggled to find a set of parameters that minimised the objective error function with the data we collected, so we constructed more data around our measurements through interpolation. Thanks to that extra data the parameter optimisation algorithm found a set of parameters that produced *mitfa* and *dct* trajectories that were very close to the experimental data.

The new model suggests that low expression values of *sox10* are sufficient to drive *mitfa* expression in high levels. It also highlights that *kit* is not required to be highly expressed either to maintain those high *mitfa* levels. The model predicts that *mitfa* expression alone is not enough to reach the high expression levels of *dct* seen in the

data. This is achieved by the high expression of *sox9b* which becomes an important player in driving *dct* expression. This prediction was unexpected considering the experimental data available which suggested that the *sox9b* dependent *dct* expression is not as important as the activation via *mitfa*. Hence the model was modified to try to accommodate this experimental observations: the *kit* equation was modified, the parameters that control the *sox9b* trajectory were left out of the optimisation, and the 20 hpf *mitfa* data point was made less important in the parameter optimisation algorithm. After these changes the model produced better trajectories for some genes and worse for some others.

The parameter optimisation procedure has highlighted that more data is needed to obtain good parameter values that can explain the data: both from more genes in the network and from more stages. This model represents an initial attempt to develop a quantitative mathematical model for the melanocyte GRN. The step forward would be to collect single cell gene expression measurements from these and more genes at the same time using a multiplex technology such as nCounter from Nanostring. In this way we would have accurate single cell data to perform a parameter optimisation procedure and produce a quantitative model of the GRN.

Chapter 6:
Final discussion

The complex genetic interactions that regulate the differentiation and specification of neural crest cells to melanocytes make these cells a very interesting and powerful model to understand differentiation and stem cell fate choice. As a multipotent stem cell differentiates, its GRN becomes reconfigured from a state where multiple fates can be chosen to a more stable state that indicates commitment to a specific cell fate. In the Waddington's epigenetic landscape the cell fate commitment is represented by a ball rolling into a valley separated by a ridge, and the different genetic interactions impose developmental constrains in the multipotent progenitor that will determine which of the available fates or valleys a particular cell chooses.

The modern picture of the Waddington's epigenetic landscape is explained better by GRN dynamics (Kauffman, 1969; Mendoza and Alvarez-Buylla, 1998; Zhou *et al.*, 2012). A cell is considered a dynamical system, and its state at a certain time can be described by a set of time-dependent variables: the amount of the different proteins within the cell or the levels of gene expression (Huang and Kauffman, 2013). Each gene in the cell's GRN representing one variable.

To understand this dynamical system concept it is useful to image an abstract space called state space, which in the GRNs context comprises all the theoretically possible states a cell can display; each point in this abstract space represents one particular expression profile. Furthermore, it is assumed that the cell state at a certain time and the cell state at a later time are connected (Velderrain *et al.*, 2015).

However, within the whole state space only a small number of stationary or quasi-stationary gene configurations will satisfy the constraints imposed by the GRN (Kauffman, 1969). When these steady states are also resilient to perturbations, that is, if they return back to the steady state after being kicked away by internal or external variations, we refer to them as attractors. All other states are either unstable

or form part of transitory trajectories that are guided toward one of these attractor states. The theory suggests that attractor states correspond to the different cell types; and that these emerge as a consequence of the dynamical constraints imposed by the underlying GRN (Huang *et al.*, 2009; Huang and Kauffman, 2013).

In this project we are trying to understand the properties and constrains of a GRN that make a cell state become an attractor, and hence a stable differentiated cell type. We are using the differentiation of melanocytes from their neural crest cells progenitors as a model to understand the properties of its GRN that produce a stable cell state.

Melanocytes, one of the neural crest derived pigment cells, are characterised by the production of melanin, a pigment molecule located in special organelles called melanosomes. They have been thoroughly studied to understand development and fate choice, but also because they are the cause of many severe human diseases (melanoma, Waardenburg syndrome or Tietz syndrome, for example). Some of the main genes driving melanocyte development have been studied for a long time and we have a good picture of some of their interactions that regulate melanocyte development. Greenhill *et al.* (2011) developed a gene regulatory network incorporating the main genes and interactions known to play a role in melanocyte biology, and generated a mathematical model to help understand this complex network and make specific predictions about its properties.

They build the GRN and mathematical model in an iterative way starting with an initial model which allowed us to make predictions on how the gene expression would behave, then they tested those predictions and modified the initial model until they got a complete enough model that was able to explain their experiments. In order to build this model they used ISH gene expression data, which meant that they have semi quantitative data at best. Therefore they were able to obtain a model that was able to describe the trends and shape of the curves of gene expression along time to match what they saw in their ISH data. Despite this model being semi-quantitative they manage to make predictions and validate them in experiments, so it proved to be a useful tool (Greenhill *et al.*, 2011).

In this project we sought to collect a dense time series gene expression data of four of the main genes of the melanocyte GRN (*sox10*, *kit*, *mitfa* and *dct*) to transform the semi-quantitative mathematical model established by Greenhill *et al.* (2011) into a quantitative model that is able to describe the data we collected. Moreover, we intended to identify more genes that are part of the melanocyte development process to be incorporated in the GRN by analysing microarray data and performing ISH experiments.

The Greenhill *et al.* (2011) GRN describes the behaviour of nine genes involved in melanocyte development, but these are not the only genes that play a role in this process. Hence, we aimed to find more genes involved in melanocyte development to incorporate into the GRN. For that we analysed the microarray data produced by Chipperfield in 2009 using a more advance and versatile approach based in R.

Microarray studies have been successfully used in other examples to identify genes related to a certain cell type or biological function in zebrafish before. Weger *et al.* (2011) used a similar setup to ours (Affymetrix array and Bioconductor analysis) to identify over 100 genes that are upregulated by light in zebrafish. They were able to identify genes involved in circadian clock function, DNA repair, retinal light reception and metabolism, as well as the enrichment of both E- and D-box elements in the promoters of the light induced genes. Many other examples have used microarray analysis comparing wild type with mutant gene expression; for example, van der Vaart *et al.* (2013) compared the expression profile of *myd88* mutant and wild type embryos and identified several transcription factors central to innate immunity that are regulated by Myd88 signalling.

In our microarray analysis and subsequent validation through in situ hybridisation, we have identified five genes as downregulated in *mitfa* mutant embryos: *atp6v1ab, atp6v1e1b, degs1, pah* and *tspan36*. These genes have a clear role in melanocyte biology, as they are involved in melanosome maturation, melanin synthesis and melanocyte migration. We suggest that these genes are placed downstream of *mitfa* in the GRN. We also used the list of downregulated genes in *sox10* mutant embryos obtained from the microarray analysis to explore the ZFIN database looking for information available on the expression of these genes. From this analysis we found

seven more genes that are likely to be part of the GRN as they have a role in melanocyte biology: *rab32a, rab7, bace2, ednrba, gart, ppat* and *atp6ap1b*. However these genes would need to be investigated in further experiments.

By using microarrays to find downregulated genes in a mutant background we identified genes that are regulated by the mutated gene (*sox10*) but we do not know if that interaction is direct or indirect. Therefore, in order to find out if the candidate genes are directly regulated by *sox10* we should perform a ChIP-seq analysis. This kind on analysis was successfully used to identify that the transcription factor FOXD3, which plays an essential role in pre-migratory neural crest development, directly binds to regions of the TWIST1 promoter, leading to its repression in human mutant BRAF melanoma cells (Weiss *et al.*, 2014). The main issue with this approach is that we need suitable anti-Sox10 antibodies, which are difficult to obtain. However, there are companies that are working on this issue. The advantage is that once we have an antibody that works with our sample we can perform the ChIP-seq experiment and identify all the genes that are directly regulated by Sox10. Despite the difficulty of obtaining suitable antibodies some studies have successfully used ChIP-seq with zebrafish embryos. Leichsenring *et al.* (2013) found that zebrafish *pou5f1*, a homolog of the mammalian pluripotency transcription factor Oct4, activates many of the first zygotically expressed genes in the zebrafish embryo. In mammalian ES cells, the Pou5f1-Sox2 complex cooperatively assembles to target early zygotic genes for activation (Loh *et al.*, 2006). To detect in vivo Pou5f1 and Sox2 chromatin binding events, Leichsenring *et al.* performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) and identified many genes regulated by these transcription factors. Their data position Pou5f1 and Sox2 transcription factors at the centre of the zygotic gene activation network of vertebrates (Leichsenring *et al.*, 2013). These studies prove how useful and powerful ChIP-seq is to identify transcriptional regulation.

Greenhill *et al.* (2011) identified though their mathematical model the need of a factor (named factor Y) in the network to maintain *mitfa* expression and allow stable melanocyte differentiation. It has been shown that the presence of positive feedback loops in gene regulatory networks increase the robustness and stability of the

system. Positive feedback loops, like the one in between mitfa and factor Y, promote the ability of a system to become activated in response to small input signal (Mitrophanov and Groisman, 2008). This means that our gene regulatory network does not need a high amount of Sox10 to drive enough mitfa expression and allow melanocyte specification. We sought to find the gene that may take the role of factor Y. We proposed that *kit* may be this factor as the literature suggested that *kit* expression might be driven by Mitfa while *kit* itself may regulate *mitfa* transcription (Hou *et al.*, 2000). We have confirmed that Mitfa drives *kit* expression, as well as shown that *kit* expression is transiently driven by Sox10 at early stages of development. As part of the feedback loop, *kit* seems to drive and maintain *mitfa* expression, as *mitfa* expression appears to be reduced in *kit* mutants. However this needs to be quantified. Therefore we will need to perform a *mitfa* qPCR in *kit* mutants and compare it with wild type fish, at several stages to confirm if lack of *kit* activity hinders *mitfa* expression, and hence if *kit* regulates *mitfa* transcription.

Qualitative models of gene regulatory networks have been used quite widely as they have two main advantages: they can be built from limited experimental observations or knowledge-based information; and they make possible to hypothesise about the dynamics of biological regulatory systems despite the lack of kinetic information at transcriptional level (Bourdon *et al.*, 2011). Although these approaches provide high-level insights into the functioning of gene networks, they do not accurately describe the real dynamics of GRNs. Hence our aim was to transform the qualitative or semi-quantitative model developed by Greenhill *et al.* (2011) into a quantitative one. For that we needed to collect data that we used to obtain the parameter values for the ordinary differential equations of the mathematical model that describes our data. We obtained an estimate of gene expression per melanocyte by measuring gene expression in 5 embryo trunks and then counting the gene expressing cells in the trunk. We extracted and measured through qPCR the gene expression of *sox10*, *kit*, *mitfa* and *dct* from 5 embryo trunks. As *sox10* and *kit* are expressed in many other cell types besides melanocytes their gene expression measurement is not only measuring melanocyte gene expression, which is what we were modelling, so it was not used. To get a measurement of the cells expressing those genes in the trunk, we

performed ISH for our four genes and counted the gene expressing cells in the trunk. We used both of these measurements qPCR copy number and cell number, to estimate the copy number per cell for *mitfa* and *dct*. This estimate was used to perform the parameter optimisation, which aims to find the best parameter values for the mathematical model so its predictions are as close as possible to the experimental data collected. For that we created a Matlab algorithm that uses an objective function to calculate the error between the predicted trajectory calculated with a new set of parameters and the measured data. Then the nonlinear least-squares solver finds new parameter values that minimise that error function.

Our algorithm struggled to fit the model to the data as despite having collected a very dense time series expression data for our genes the number of parameters to fit exceeded by far the amount of data available. So we constructed more data around our measurements through interpolation. Thanks to that extra data the parameter optimisation algorithm found a set of parameters that produced *mitfa* and *dct* trajectories that were very close to the experimental data.

Davidson formulated the concept of developmental gene regulatory networks, and his laboratory has been investigating the GRN that controls the development of sea urchin embryos for many years. They have built a GRN with over 60 regulatory and signalling genes that regulates the specification of the endomesoderm in sea urchin embryos (Smith *et al.*, 2008). They established a predictive, dynamic Boolean model that describes spatial and temporal gene expression according to the regulatory logic and gene interactions specified in the GRN for embryonic development in the sea urchin. Additional information that was used as input into this model included the embryonic geometry and gene expression kinetics (Bolouri and Davidson, 2003). The resulting model predicted gene expression patterns for a large number of genes each hour up to gastrulation in different spatial domains of the embryo. Despite being a Boolean model that only takes into account if an interaction is acting or not, a direct comparison with experimental observations showed that the model predicted the gene expression patterns with notable spatial and temporal accuracy, proving that even simpler Boolean models are sufficient to explain complex developmental process (Peter *et al.*, 2012). Davidson's GRN is much more complex in term of number

of genes than the melanocyte GRN, however their mathematical model is simpler that the one we are aiming to produce, as we aim to develop a dynamical model that describes the rate of change of gene expression over time.

An example of a dynamical model whose set up is very similar to ours is the development of a model for the regulation of flowering time in *Arabidopsis thaliana* by Valentim *et al.* (2015). Model parameters were estimated based on expression time-courses for relevant genes obtained by qPCR from leaves. They modelled 6 genes, which is more similar to the number of genes in the melanocyte GRN than to the Davidson's model. They used Hill equations to model the kinetics of the gene expression like I did, and used ordinary differential equations to describe the interactions between the network components. Like us they assumed that the relationship between RNA and protein levels is linear. They used a Matlab algorithm to fit the model to the data and estimated the model parameters using the *lsqnonlin* solver as I did. Their model predicts that a disturbance in a particular gene has not necessarily a big impact on directly connected genes. They also predicted some other more specific features of their GNR, such as the importance of cooperativity in the regulation of APETALA1 by LFY. Lastly, they validated the model by comparing predicted expression time-courses for mutants in components of the network with experimental data (Valentim *et al.*, 2015). Another example of dynamical models is the one produced by Aguilar-Hidalgo *et al.* (2013) of the GRN underlying the *Drosophila* ocellar complex. Unlike ours, their model distinguish between mRNA transcription and translation. However, they also use Hill equations to model the changes of gene expression in time, like I did. Their model comprises 13 ordinary differential equations to describe the behaviour of each one of the system variables. Unlike us, or Valentim *et al.* (2015) that used qPCR expression data to test the model, they used immunostaining and quantification of those images to validate the model. The mathematical model provided them a better understanding of how the hedgehog signalling gradient and the gene regulatory network interact to generation a fully developed ocellar complex (Aguilar-Hidalgo *et al.*, 2013). Therefore our model has been produced using similar methos to the ones used successfully bu other laboratories.

The new model obtained after the parameter optimisation procedure suggests that low expression values of *sox10* are sufficient to drive *mitfa* expression in high levels. It also highlights that *kit* is not required to be highly expressed either to maintain those high *mitfa* levels. The model predicts that *mitfa* expression alone is not enough to reach the high expression levels of *dct* seen in the data. This is achieved by the high expression of *sox9b* which becomes an important player in driving *dct* expression. This prediction was unexpected considering the experimental data available which suggested that the *sox9b* dependent *dct* expression is not as important as the activation via *mitfa*. We modified the model to improve the equation that regulate *kit* expression, exclude the *sox9b* parameters from the optimisation so this gene is not used as the main driver of *dct* expression, and exclude the first *mitfa* data point as it seems to be an outlier. These changes improve the trajectories for some genes such as *sox10*, *mitfa* and *kit*; however the trajectories for some other genes (*dct* or *tyrp1*) are still very far from the experimental data. In order to improve our mathematical model we would need to obtain more data, ideally from more genes and more stages. In that way the parameter optimisation algorithm will be able to find better values to produce a model that explains more accurately the data. We could also use other parameter optimisation methods available such as the Monter Carlo method. Eventually the long term objective is to measure gene expression in single cells. For that we could use FACS to sort single cells and a technology like nCounter from NanoString that allows to measure hundreds of genes at the same time even in diluted DNA samples coming from single cells. The NanoString nCounter technology was used by Sandler and Stathopoulos (2016) to study maternal-to-zygotic transition in *Drosophila melanogaster* embryos. This event marks the rapid turning point when zygotic transcription begins and control of development is transferred from maternal transcripts. Characterising the sequential activation of the genome during the MZT requires precise timing and a sensitive assay to measure changes in expression. They used the nCounter technology, which directly counts mRNA transcripts without reverse transcription or amplification, to study more than 70 genes expressed early in *Drosophila* embryos. The study provided new insights into how the GRN players are sequentially activated in the maternal-to-zygotic transition and suggested that network properties regulate levels of

transcription for groups of genes (Sandler and Stathopoulos, 2016). This study demonstrates that the NanoString technology can be used to precisely quantify transcripts and create a fine-scale time course of gene expression along embryonic development, and hence it would be a very powerful technology to use in our research.

There have been some efforts in building gene regulatory networks to understand neural crest development in zebrafish before, such as the studies carried out by Van Otterloo and Wang, which established the interactions between the transcriptions factors, Tfap2a, Foxd3, Bmp, Wnt and Sox10 to regulate neural crest induction (Van Otterloo *et al.*, 2012; Wang *et al.*, 2011). The Pavan laboratory has also put a lot of effort in building the GRN that describes the genes involved in the regulation of melanocytes and RPE cells in mice (Gorkin *et al.*, 2012; Fufa *et al.*, 2015). With their own genomic data obtained through ChIP-seq, microarrays and qPCR, and the data they compiled from other sources they have built one of the most comprehensive GRNs for pigment cells. However, despite all the knowledge about the genes involved in melanocyte development, ours is one of the first and major attempts to produce a mathematical modelling of the GRN that aims to get a better understanding of the emergent properties of the gene network.

To sum up in this project we have identified several genes that play a role in melanocyte biology and have included some of them in the GRN. The next step would be to develop the equations in the mathematical model to describe them. We have identified *kit* as Factor Y in the Greenhill *et al.* (2011) model, although we need further testing to assess some of the details of its interaction with *mitfa*. Finally we have obtained quantitative gene expression data from four genes and 12 stages along development that have allowed us to start the first steps into the parameter optimisation of the model towards developing a fully quantitative mathematical model. The modelling has allowed us to make some predictions that would need to be experimentally tested in the future. Further work is needed, but this study represents the first attempt to obtain densely populated gene expression data from melanocytes for the development of a quantitative mathematical model in zebrafish.

References

Adameyko, I., Lallemend, F., Aquino, J. B., Pereira, J. A., Topilko, P., Müller, T., Fritz, N., Beljajeva, A., Mochii, M., Liste, I., Usoskin, D., Suter, U., Birchmeier, C. & Ernfors, P. (2009) Schwann cell precursors from nerve innervation are a cellular origin of melanocytes in skin. *Cell*. 139 (2), 366–379.

Aguilar-Hidalgo, D., Domínguez-Cejudo, M. A., Amore, G., Brockmann, A., Lemos, M. C., Córdoba, A. & Casares, F. (2013) A Hh-driven gene network controls specification, pattern and size of the Drosophila simple eyes. *Development (Cambridge, England)*. 140 (1), 82–92.

Alcasabas, A. A., de Clare, M., Pir, P. & Oliver, S. G. (2013) Control analysis of the eukaryotic cell cycle using gene copy-number series in yeast tetraploids. *BMC genomics*. 14 (1), 744.

Alexeev, V. & Yoon, K. (2006) Distinctive role of the cKit receptor tyrosine kinase signaling in mammalian melanocytes. *Journal of Investigative Dermatology*. [Online] 126 (5), 1102–1110.

Alonso-Curbelo, D., Osterloh, L., Cañón, E., Calvo, T. G., Martínez-Herranz, R., Karras, P., Martínez, S., Riveiro-Falkenbach, E., Romero, P.-O., Rodríguez-Peralto, J. L., Pastor, J. & Soengas, M. S. (2015) RAB7 counteracts PI3K-driven macropinocytosis activated at early stages of melanoma development. Oncotarget 6 (14) p.11848–11862.

Andersen, Ø., Johnsen, H., De Rosa, M. C., Præbel, K., Stjelja, S., Kirubakaran, T. G., Pirolli, D., Jentoft, S. & Fevolden, S.-E. (2015) Evolutionary history and adaptive significance of the polymorphic Pan I in migratory and stationary populations of Atlantic cod (Gadus morhua). *Marine genomics*. 2245–54.

Bagnara, J. T. & Matsumoto, J. (2007) 'Comparative Anatomy and Physiology of Pigment Cells in Nonmammalian Tissues', in JJ Nordlund et al. (eds.) *The Pigmentary System Physiology and Pathophysiology*. Second Edi Oxford, UK: Blackwell Publishing Ltd. pp. 307–344.

Van Bebber, F., Hruscha, A., Willem, M., Schmid, B. & Haass, C. (2013) Loss of Bace2 in zebrafish affects melanocyte migration and is distinct from Bace1 knock out phenotypes. *Journal of neurochemistry*. 127 (4), 471–481.

Bengtsson, M., Hemberg, M., Rorsman, P. & Ståhlberg, A. (2008) Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. *BMC Molecular Biology*. 9 (1), 63.

Bertolotto, C., Abbe, P., Hemesath, T. J., Bille, K., Fisher, D. E., Ortonne, J. P. & Ballotti, R. (1998) Microphthalmia gene product as a signal transducer in cAMP-induced differentiation of melanocytes. *The Journal of cell biology*. [Online] 142 (3), 827–835.

Billon, N., Iannarelli, P., Monteiro, M. C., Glavieux-Pardanaud, C., Richardson, W. D., Kessaris, N., Dani, C. & Dupin, E. (2007) The generation of adipocytes by the neural crest. *Development (Cambridge, England)*. 134 (12), 2283–2292.

Le Bin, G. C., Muñoz-Descalzo, S., Kurowski, A., Leitch, H., Lou, X., Mansfield, W., Etienne-Dumeau, C., Grabole, N., Mulas, C., Niwa, H., Hadjantonakis, A.-K. & Nichols, J. (2014) Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development (Cambridge, England)*. 141 (5), 1001–1010.

Bolouri, H. & Davidson, E. H. (2003) Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proceedings of the National Academy of Sciences of the United States of America*. 100 (16), 9371–9376.

Bondurand, N., Pingault, V., Goerich, D. E., Lemort, N., Sock, E., Le Caignec, C., Wegner, M. & Goossens, M. (2000) Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Human molecular genetics*. [Online] 9 (13), 1907–1917.

Bourdon, J., Eveillard, D. & Siegel, A. (2011) Integrating quantitative knowledge into a qualitative gene regulatory network. *PLoS computational biology*. 7 (9), e1002157.

Brennecke, P., Anders, S., Kim, J. K., Kołodziejczyk, A. A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S. A., Marioni, J. C. & Heisler, M. G. (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nature methods*. 10 (11), 1093–1095.

Bronner-Fraser, M. (2002) Molecular analysis of neural crest formation. *Journal of Physiology-Paris*. 96 (1-2), 3–8.

Bronner-Fraser, M. (1995) Origins and Developmental Potential of the Neural Crest. *Experimental Cell Research*. 218 (2), 405–417.

Bustin, S. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*. 25 (2), 169–193.

Caporali, S., Alvino, E., Levati, L., Esposito, A. I., Ciomei, M., Brasca, M. G., Del Bufalo, D., Desideri, M., Bonmassar, E., Pfeffer, U. & D'Atri, S. (2012) Down-regulation of the PTTG1 proto-oncogene contributes to the melanoma

suppressive effects of the cyclin-dependent kinase inhibitor PHA-848125. *Biochemical pharmacology*. 84 (5), 598–611.

Carey, M. F., Peterson, C. L. & Smale, S. T. (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harbor protocols*. 2009 (9), pdb.prot5279.

Carney, T. J., Dutton, K. A., Greenhill, E., Delfino-Machín, M., Dufourcq, P., Blader, P. & Kelsh, R. N. (2006) A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development (Cambridge, England)*. 133 (23), 4619–4630.

Carreira, S., Goodall, J., Aksan, I., La Rocca, S. A., Galibert, M.-D., Denat, L., Larue, L. & Goding, C. R. (2005) Mitf cooperates with Rb1 and activates p21Cip1 expression to regulate cell cycle progression. *Nature*. [Online] 433 (7027), 764–769.

Chipperfield, T. R. (2009) *Analysis of Sox10 Target Genes in Zebrafish Early Development*. [Online]. University of Bath.

Cooper, C. D., Linbo, T. H. & Raible, D. W. (2009) Kit and foxd3 genetically interact to regulate melanophore survival in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*. 238 (4), 875–886.

Cui, X. & Churchill, G. (2003) Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology*. 4 (4), 210.

Curran, K., Lister, J. A., Kunkel, G. R., Prendergast, A., Parichy, D. M. & Raible, D. W. (2010a) Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafish neural crest. *Developmental Biology*. [Online] 344 (1), 107–118.

Curran, K., Lister, J. A., Kunkel, G. R., Prendergast, A., Parichy, D. M. & Raible, D. W. (2010b) Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafish neural crest. *Developmental biology*. 344 (1), 107–118.

Curran, K., Raible, D. W. & Lister, J. A. (2009) Foxd3 controls melanophore specification in the zebrafish neural crest by regulation of Mitf. *Developmental biology*. 332 (2), 408–417.

Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Schilstra, M. J., Clarke, P. J. C., Rust, A. G., Pan, Z., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L. & Bolouri, H. (2002) A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental biology*. 246 (1), 162–190.

Donoghue, P. C. J., Graham, A. & Kelsh, R. N. (2008) The origin and evolution of the neural crest. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 30 (6), 530–541.

Dorsky, R. I., Moon, R. T. & Raible, D. W. (1998) Control of neural crest cell fate by the Wnt signalling pathway. *Nature*. 396 (6709), 370–373.

Dorsky, R. I., Raible, D. W. & Moon, R. T. (2000) Direct regulation of nacre, a zebrafish MITF homolog required for pigment cell formation, by the Wnt pathway. *Genes & Development*. [Online] 14 (2), 158–162.

Dutton, J. R., Antonellis, A., Carney, T. J., Rodrigues, F. S. L. M., Pavan, W. J., Ward, A. & Kelsh, R. N. (2008) An evolutionarily conserved intronic region controls the spatiotemporal expression of the transcription factor Sox10. *BMC developmental biology*. 8105.

Dutton, K. a, Pauliny, A., Lopes, S. S., Elworthy, S., Carney, T. J., Rauch, J., Geisler, R., Haffter, P. & Kelsh, R. N. (2001) Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development*. [Online] 128 (21), 4113–4125.

Dutton, K., Pauliny, A., Lopes, S. S., Elworthy, S., Carney, T. J., Rauch, J., Geisler, R., Haffter, P. & Kelsh, R. N. (2001) Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development (Cambridge, England)*. [Online] 128 (21), 4113–4125.

Ellis, K., Bagwell, J. & Bagnat, M. (2013) Notochord vacuoles are lysosome-related organelles that function in axis and spine morphogenesis. *The Journal of cell biology*. 200 (5), 667–679.

Elworthy, S. (2003) Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development*. 130 (12), 2809–2818.

Elworthy, S., Lister, J. A., Carney, T. J., Raible, D. W. & Kelsh, R. N. (2003) Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development (Cambridge, England)*. [Online] 130 (12), 2809–2818.

Ferrell, J. E. (2012) Bistability, bifurcations, and Waddington's epigenetic landscape. *Current biology : CB*. 22 (11), R458–R466.

Fufa, T. D., Harris, M. L., Watkins-Chow, D. E., Levy, D., Gorkin, D. U., Gildea, D. E., Song, L., Safi, A., Crawford, G. E., Sviderskaya, E. V, Bennett, D. C., Mccallion, A. S., Loftus, S. K. & Pavan, W. J. (2015) Genomic analysis reveals distinct mechanisms and functional classes of SOX10-regulated genes in melanocytes. *Human molecular genetics*. 24 (19), 5433–5450.

Fukuda, M., Kuroda, T. S. & Mikoshiba, K. (2002) Slac2-a/melanophilin, the missing link between Rab27 and myosin Va: implications of a tripartite protein complex for melanosome transport. *The Journal of Biological Chemistry*. [Online] 277 (14), 12432–12436.

Furusawa, C. & Kaneko, K. (2012) A dynamical-systems view of stem cell biology. *Science (New York, N.Y.)*. 338 (6104), 215–217.

Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T. & Hopkins, N. (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature*. [Online] 383 (6603), 829–832.

Gammill, L. S. & Bronner-Fraser, M. (2003) Neural crest specification: migrating into genomics. *Nature reviews. Neuroscience*. 4 (10), 795–805.

García-Borrón, J. C., Sánchez-Laorden, B. L. & Jiménez-Cervantes, C. (2005) Melanocortin-1 receptor structure and functional regulation. *Pigment cell research*. [Online] 18 (6), 393–410.

García-Castro, M. I., Marcelle, C. & Bronner-Fraser, M. (2002) Ectodermal Wnt function as a neural crest inducer. *Science (New York, N.Y.)*. [Online] 297 (5582), 848–851.

Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. (2004) affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*. 20 (3), 307–315.

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., Hood, L. & Dimitrov, K. (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology*. 26 (3), 317–325.

Gilbert, S. F. (2000) *Developmental Biology*. 6th edition. [Online]. Sunderland (MA): Sinauer Associates.

Giovannetti, E., Wang, Q., Avan, A., Funel, N., Lagerweij, T., Lee, J.-H., Caretti, V., van der Velde, A., Boggi, U., Wang, Y., Vasile, E., Peters, G. J., Wurdinger, T. & Giaccone, G. (2014) Role of CYB5A in pancreatic cancer prognosis and autophagy modulation. *Journal of the National Cancer Institute*. 106 (1), djt346.

Goda, M. & Fujii, R. (1995) Blue Chromatophores in Two Species of Callionymid Fish. *Zoological Science*. [Online] 12 (6), 811–813.

Goda, M., Ohata, M., Ikoma, H., Fujiyoshi, Y., Sugimoto, M. & Fujii, R. (2011) Integumental reddish-violet coloration owing to novel dichromatic

chromatophores in the teleost fish, Pseudochromis diadema. *Pigment cell & melanoma research*. [Online] 24 (4), 614–617.

Goding, C. R. (2000) Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. *Genes & Development*. [Online] 141712–1728 .

Goodrich, H. B., Hill, G. A. & Arrick, M. S. (1941) The Chemical Identification of Gene-Controlled Pigments in Platypoecilus and Xiphophorus and Comparisons with Other Tropical Fish. *Genetics*. [Online] 26 (6), 573–586.

Gorkin, D. U., Lee, D., Reed, X., Fletez-Brant, C., Bessling, S. L., Loftus, S. K., Beer, M. A., Pavan, W. J. & McCallion, A. S. (2012) Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome research*. gr.139360.112 – .

Greenhill, E. R., Rocco, A., Vibert, L., Nikaido, M. & Kelsh, R. N. (2011) An Iterative Genetic and Dynamical Modelling Approach Identifies Novel Features of the Gene Regulatory Network Underlying Melanocyte Development Mary C. Mullins (ed.). *PLoS Genetics*. 7 (9), e1002265.

Guillot, R., Ceinos, R. M., Cal, R., Rotllant, J. & Cerdá-Reverter, J. M. (2012) Transient ectopic overexpression of agouti-signalling protein 1 (asip1) induces pigment anomalies in flatfish. *PloS one*. [Online] 7 (12), e48526.

Le Guyader, S., Maier, J. & Jesuthasan, S. (2005) Esrom, an ortholog of PAM (protein associated with c-myc), regulates pteridine synthesis in the zebrafish. *Developmental biology*. 277 (2), 378–386.

Hallsson, J. H., Haflidadóttir, B. S., Stivers, C., Odenwald, W., Arnheiter, H., Pignoni, F. & Steingrímsson, E. (2004) The basic helix-loop-helix leucine zipper transcription factor Mitf is conserved in Drosophila and functions in eye development. *Genetics*. [Online] 167 (1), 233–241.

Hearing, V. J. (2011) Determination of melanin synthetic pathways. *The Journal of investigative dermatology*. 131 (E1), E8–E11.

Hearing, V. J. & Jiménez, M. (1987) Mammalian tyrosinase--the critical regulatory control point in melanocyte pigmentation. *The International journal of biochemistry*. [Online] 19 (12), 1141–1147.

Henion, P. D., Raible, D. W., Beattie, C. E., Stoesser, K. L., Weston, J. A. & Eisen, J. S. (1996) Screen for mutations affecting development of Zebrafish neural crest. *Developmental Genetics*. [Online] 18 (1), 11–17.

Hida, T., Sohma, H., Kokai, Y., Kawakami, A., Hirosaki, K., Okura, M., Tosa, N., Yamashita, T. & Jimbow, K. (2011) Rab7 is a critical mediator in vesicular

transport of tyrosinase-related protein 1 in melanocytes. *The Journal of dermatology*. 38 (5), 432–441.

Higdon, C. W., Mitra, R. D. & Johnson, S. L. (2013) Gene expression analysis of zebrafish melanocytes, iridophores, and retinal pigmented epithelium reveals indicators of biological function and developmental origin. *PloS one*. 8 (7), e67801.

Hoashi, T., Sato, S., Yamaguchi, Y., Passeron, T., Tamaki, K. & Hearing, V. J. (2010) Glycoprotein nonmetastatic melanoma protein b, a melanocytic cell marker, is a melanosome-specific and proteolytically released protein. *The FASEB journal*. [Online] 24 (5), 1616–1629.

Hodgkinson, C. A., Moore, K. J., Nakayama, A., Steingrímsson, E., Copeland, N. G., Jenkins, N. A. & Arnheiter, H. (1993) Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell*. [Online] 74 (2), 395–404.

Hou, L., Arnheiter, H. & Pavan, W. J. (2006) Interspecies difference in the regulation of melanocyte development by SOX10 and MITF. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 103 (24), 9081–9085.

Hou, L., Panthier, J. J. & Arnheiter, H. (2000) Signaling and transcriptional regulation in the neural crest-derived melanocyte lineage: interactions between KIT and MITF. *Development (Cambridge, England)*. [Online] 127 (24), 5379–5389.

Hou, L. & Pavan, W. J. (2008) Transcriptional and signaling regulation in neural crest stem cell-derived melanocyte development: do all roads lead to Mitf? *Cell research*. 18 (12), 1163–1176.

Howe, D. G., Bradford, Y. M., Conlin, T., Eagle, A. E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S. A. T., Paddock, H., Pich, C., Ramachandran, S., Ruef, B. J., Ruzicka, L., Schaper, K., Shao, X., Singer, A., Sprunger, B., Van Slyke, C. E. & Westerfield, M. (2013) ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. *Nucleic acids research*. 41 (Database issue), D854–D860.

Huang, S., Ernberg, I. & Kauffman, S. (2009) Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in cell & developmental biology*. 20 (7), 869–876.

Huang, S. & Kauffman, S. (2013) How to escape the cancer attractor: rationale and limitations of multi-target drugs. *Seminars in cancer biology*. 23 (4), 270–278.

Hume, A. N. & Seabra, M. C. (2011) Melanosomes on the move: a model to understand organelle dynamics. *Biochemical Society transactions*. [Online] 39 (5), 1191–1196.

Ichikawa, Y., Ohtani, H. & Miura, I. (1998) The erythrophore in the larval and adult dorsal skin of the brown frog, Rana ornativentris: its differentiation, migration, and pigmentary organelle formation. *Pigment cell research*. [Online] 11 (6), 345–354.

Ignatius, M. S., Moose, H. E., El-Hodiri, H. M. & Henion, P. D. (2008) colgate/hdac1 Repression of foxd3 expression is required to permit mitfa-dependent melanogenesis. *Developmental Biology*. [Online] 313 (2), 568–583.

Inoue, S., Kondo, S., Parichy, D. M. & Watanabe, M. (2014) Tetraspanin 3c requirement for pigment cell interactions and boundary formation in zebrafish adult pigment stripes. *Pigment Cell & Melanoma Research*. 27 (2), 190–200.

Ito, S. & Wakamatsu, K. (2003) Quantitative analysis of eumelanin and pheomelanin in humans, mice, and other animals: a comparative review. *Pigment cell research*. [Online] 16 (5), 523–531.

Jiao, Z., Mollaaghababa, R., Pavan, W. J., Antonellis, A., Green, E. D. & Hornyak, T. J. (2004) Direct interaction of Sox10 with the promoter of murine Dopachrome Tautomerase (Dct) and synergistic activation of Dct expression with Mitf. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society*. 17 (4), 352–362.

Johnson, S. L., Africa, D., Walker, C. & Weston, J. A. (1995) Genetic control of adult pigment stripe development in zebrafish. *Developmental Biology*. [Online] 167 (1), 27–33.

Johnson, S. L., Nguyen, A. N. & Lister, J. A. (2011a) mitfa is required at multiple stages of melanocyte differentiation but not to establish the melanocyte stem cell. *Developmental Biology*. [Online] 350 (2), 405–413.

Johnson, S. L., Nguyen, A. N. & Lister, J. A. (2011b) mitfa is required at multiple stages of melanocyte differentiation but not to establish the melanocyte stem cell. *Developmental biology*. 350 (2), 405–413.

Jordan, S. A. & Jackson, I. J. (2000) MGF (KIT ligand) is a chemokinetic factor for melanoblast migration into hair follicles. *Developmental Biology*. [Online] 225 (2), 424–436.

Karlsson, J., von Hofsten, J. & Olsson, P. E. (2001) Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. *Marine biotechnology (New York, N.Y.)*. 3 (6), 522–527.

Kauffman, S. (1969) Homeostasis and differentiation in random genetic control networks. *Nature*. [Online] 224 (5215), 177–178.

Kedra, D., Pan, H.-Q., Seroussi, E., Fransson, I., Guilbaud, C., Collins, J. E., Dunham, I., Blennow, E., Roe, B. A., Piehl, F. & Dumanski, J. P. (1998) Characterization of the human synaptogyrin gene family. *Human Genetics*. 103 (2), 131–141.

Kelsh, R. N. (2006a) Sorting out Sox10 functions in neural crest development. *BioEssays*. 28 (8), 788–798.

Kelsh, R. N. (2006b) Sorting out Sox10 functions in neural crest development. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 28 (8), 788–798.

Kelsh, R. N., Brand, M., Jiang, Y. J., Heisenberg, C. P., Lin, S., Haffter, P., Odenthal, J., Mullins, M. C., Van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Kane, D. A., Warga, R. M., Beuchle, D., Vogelsang, L. & Nüsslein-Volhard, C. (1996) Zebrafish pigmentation mutations and the processes of neural crest development. *Development*. [Online] 123 (1), 369–389.

Kelsh, R. N., Schmid, B. & Eisen, J. S. (2000a) Genetic analysis of melanophore development in zebrafish embryos. *Developmental Biology*. [Online] 225 (2), 277–293.

Kelsh, R. N., Schmid, B. & Eisen, J. S. (2000b) Genetic analysis of melanophore development in zebrafish embryos. *Developmental biology*. 225 (2), 277–293.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. (1995) Stages of Embryonic Development of the Zebrafish. *World Wide Web Internet And Web Information Systems*. [Online] 10.

Kippenberger, S., Bernd, A., Bereiter-Hahn, J., Ramirez-Bosca, A. & Kaufmann, R. (1998) The Mechanism of Melanocyte Dendrite Formation: The Impact of Differentiating Keratinocytes. *Pigment Cell Research*. 11 (1), 34–37.

Kristiansson, E., Thorsen, M., Tamás, M. J. & Nerman, O. (2009) Evolutionary forces act on promoter length: identification of enriched cis-regulatory elements. *Molecular biology and evolution*. 26 (6), 1299–1307.

Kühn, C., Wierling, C., Kühn, A., Klipp, E., Panopoulou, G., Lehrach, H. & Poustka, A. J. (2009) Monte Carlo analysis of an ODE Model of the Sea Urchin Endomesoderm Network. *BMC systems biology*. 3 (1), 83.

Lacour, J. P., Gordon, P. R., Eller, M., Bhawan, J. & Gilchrest, B. A. (1992) Cytoskeletal events underlying dendrite formation by cultured pigment cells. *Journal of Cellular Physiology*. [Online] 151 (2), 287–299.

Lahav, R., Ziller, C., Dupin, E. & Le Douarin, N. M. (1996) Endothelin 3 promotes neural crest cell proliferation and mediates a vast increase in melanocyte

number in culture. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 93 (9), 3892–3897.

Lamoreux, M. L., Delmas, V., Larue, L. & Bennett, D. (2010) *The Colors of Mice: A Model Genetic Network (Google eBook)*. [Online]. John Wiley & Sons.

Lang, D., Lu, M. M., Huang, L., Engleka, K. A., Zhang, M., Chu, E. Y., Lipner, S., Skoultchi, A., Millar, S. E. & Epstein, J. A. (2005) Pax3 functions at a nodal point in melanocyte stem cell differentiation. *Nature*. 433 (7028), 884–887.

Lapedriza, A., Petratou, K. & Kelsh, R. N. (2014) *Neural Crest Cells*. Elsevier.

Leal Valentim, F., Mourik, S. van, Posé, D., Kim, M. C., Schmid, M., van Ham, R. C. H. J., Busscher, M., Sanchez-Perez, G. F., Molenaar, J., Angenent, G. C., Immink, R. G. H. & van Dijk, A. D. J. (2015) A quantitative and dynamic model of the Arabidopsis flowering time gene regulatory network. *PloS one*. 10 (2), e0116973.

Lecoin, L., Sakurai, T., Ngo, M.-T., Abe, Y., Yanagisawa, M. & Le Douarin, N. M. (1998) Cloning and characterization of a novel endothelin receptor subtype in the avian class. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 95 (6), 3024–3029.

De Leeuw, S. M., Smit, N. P. M., Van Veldhoven, M., Pennings, E. M., Pavel, S., Simons, J. W. I. M. & Schothorst, A. A. (2001) Melanin content of cultured human melanocytes and UV-induced cytotoxicity. *Journal of Photochemistry and Photobiology B: Biology*. [Online] 61 (3), 106–113.

Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. (2013) Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science (New York, N.Y.)*. 341 (6149), 1005–1009.

Levine, M. & Davidson, E. H. (2005) Gene regulatory networks for development. *Proceedings of the National Academy of Sciences of the United States of America*. 102 (14), 4936–4942.

Levy, C., Khaled, M. & Fisher, D. E. (2006) MITF: master regulator of melanocyte development and melanoma oncogene. *Trends in molecular medicine*. [Online] 12 (9), 406–414.

Liem, K. F., Tremml, G. & Jessell, T. M. (1997) A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. *Cell*. [Online] 91 (1), 127–138.

Lin, J. Y. & Fisher, D. E. (2007) Melanocyte biology and skin pigmentation. *Nature*. 445 (7130), 843–850.

Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L. & Raible, D. W. (1999) nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development*. [Online] 126 (17), 3757–3767.

Lister, Robertson, C. P., Lepage, T., Johnson, S. L. & Raible, D. W. (1999) nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development (Cambridge, England)*. [Online] 126 (17), 3757–3767.

Logan, D. W., Burn, S. F. & Jackson, I. J. (2006) Regulation of pigmentation in zebrafish melanophores. *Pigment cell research*. [Online] 19 (3), 206–213.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K.-Y., Sung, K. W., Lee, C. W. H., Zhao, X.-D., Chiu, K.-P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C.-L., Ruan, Y., Lim, B. & Ng, H.-H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature genetics*. 38 (4), 431–440.

Lou, X., Kang, M., Xenopoulos, P., Muñoz-Descalzo, S. & Hadjantonakis, A.-K. (2014) A rapid and efficient 2D/3D nuclear segmentation method for analysis of early mouse embryo and stem cell image data. *Stem cell reports*. 2 (3), 382–397.

Lowery, L. A. & Sive, H. (2004) Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. *Mechanisms of development*. 121 (10), 1189–1197.

Ludwig, A., Rehberg, S. & Wegner, M. (2004) Melanocyte-specific expression of dopachrome tautomerase is dependent on synergistic gene activation by the Sox10 and Mitf transcription factors. *FEBS letters*. [Online] 556 (1-3), 236–244.

Luger, T. A., Scholzen, T. & Grabbe, S. (1997) The role of alpha-melanocyte-stimulating hormone in cutaneous biology. *The journal of investigative dermatology. Symposium proceedings / the Society for Investigative Dermatology, Inc. [and] European Society for Dermatological Research*. [Online] 2 (1), 87–93.

Marles, L. K., Peters, E. M., Tobin, D. J., Hibberts, N. A. & Schallreuter, K. U. (2003) Tyrosine hydroxylase isoenzyme I is present in human melanosomes: a possible novel function in pigmentation. *Experimental dermatology*. [Online] 12 (1), 61–70.

Marshall, W. S. (1976) Effects of hypophysectomy and ovine prolactin on the epithelial mucus-secreting cells of the Pacific staghorn sculpin, Leptocottus armatus (Teleostei: Cottidae). *Canadian journal of zoology*. [Online] 54 (10), 1604–1609.

Matesic, L. E., Yip, R., Reuss, A. E., Swing, D. A., O'Sullivan, T. N., Fletcher, C. F., Copeland, N. G. & Jenkins, N. A. (2001) Mutations in Mlph, encoding a member

of the Rab effector family, cause the melanosome transport defects observed in leaden mice. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 98 (18), 10238–10243.

Matsumoto, J. (1965) Studies on fine structure and cytochemical properties of erythrophores in swordtail, Xiphophorus helleri, with special reference to their pigment granules (Pterinosomes). *The Journal of Cell Biology*. [Online] 27 (3), 493–504.

McGill, G. G., Horstmann, M., Widlund, H. R., Du, J., Motyckova, G., Nishimura, E. K., Lin, Y.-L., Ramaswamy, S., Avery, W., Ding, H.-F., Jordan, S. A., Jackson, I. J., Korsmeyer, S. J., Golub, T. R. & Fisher, D. E. (2002) Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability. *Cell*. [Online] 109 (6), 707–718.

Mellgren, E. M. & Johnson, S. L. (2004) A requirement for kit in embryonic zebrafish melanocyte differentiation is revealed by melanoblast delay. *Development genes and evolution*. 214 (10), 493–502.

Mellgren, E. M. & Johnson, S. L. (2002) The evolution of morphological complexity in zebrafish stripes. *Trends in Genetics*. [Online] 18 (3), 128–134.

Mendoza, L. & Alvarez-Buylla, E. R. (1998) Dynamics of the genetic regulatory network for Arabidopsis thaliana flower morphogenesis. *Journal of theoretical biology*. 193 (2), 307–319.

Mitrophanov, A. Y. & Groisman, E. A. (2008) Positive feedback in cellular control systems. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 30 (6), 542–555.

Monteiro, J., Aires, R., Becker, J. D., Jacinto, A., Certal, A. C. & Rodríguez-León, J. (2014) V-ATPase proton pumping activity is required for adult zebrafish appendage regeneration. *PloS one*. 9 (3), e92594.

Montero-Balaguer, M., Lang, M. R., Sachdev, S. W., Knappmeyer, C., Stewart, R. A., De La Guardia, A., Hatzopoulos, A. K. & Knapik, E. W. (2006) The mother superior mutation ablates foxd3 activity in neural crest progenitor cells and depletes neural crest derivatives in zebrafish. *Developmental dynamics : an official publication of the American Association of Anatomists*. 235 (12), 3199–3212.

Morrison, R. L. (1995) A transmission electron microscopic (TEM) method for determining structural colors reflected by lizard iridophores. *Pigment cell research*. [Online] 8 (1), 28–36.

Murakami, T., Hijikata, T., Matsukawa, M., Ishikawa, H. & Yorifuji, H. (2006) Zebrafish protocadherin 10 is involved in paraxial mesoderm development and

somitogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists*. 235 (2), 506–514.

Ng, A., Uribe, R. A., Yieh, L., Nuckels, R. & Gross, J. M. (2009) Zebrafish mutations in gart and paics identify crucial roles for de novo purine synthesis in vertebrate pigmentation and ocular development. *Development (Cambridge, England)*. 136 (15), 2601–2611.

Nguyen, C. T., Langenbacher, A., Hsieh, M. & Chen, J.-N. (2010) The PAF1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish. *Developmental Biology*. 341 (1), 167–175.

Nishikawa, S., Kusakabe, M., Yoshinaga, K., Ogawa, M., Hayashi, S., Kunisada, T., Era, T. & Sakakura, T. (1991) In utero manipulation of coat color formation by a monoclonal anti-c-kit antibody: two distinct waves of c-kit-dependency during melanocyte development. *The EMBO journal*. [Online] 10 (8), 2111–2118.

Nitzan, E., Pfaltzgraff, E. R., Labosky, P. A. & Kalcheim, C. (2013) Neural crest and Schwann cell progenitor-derived melanocytes are two spatially segregated populations similarly regulated by Foxd3. *Proceedings of the National Academy of Sciences of the United States of America*. [Online] 110 (31), 12709–12714.

Nolan, T., Hands, R. E. & Bustin, S. A. (2006) Quantification of mRNA using real-time RT-PCR. *Nature protocols*. 1 (3), 1559–1582.

Odenthal, J., Rossnagel, K., Haffter, P., Kelsh, R. N., Vogelsang, E., Brand, M., Van Eeden, F. J., Furutani-Seiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A., Mullins, M. C. & Nüsslein-Volhard, C. (1996) Mutations affecting xanthophore pigmentation in the zebrafish, Danio rerio. *Development*. [Online] 123 (Sp. Iss. SI), 391–398.

Opdecamp, K., Kos, L., Arnheiter, H. & Pavan, W. J. (1998) Endothelin signalling in the development of neural crest-derived melanocytes. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. [Online] 76 (6), 1093–1099.

Opdecamp, K., Nakayama, A., Nguyen, M. T., Hodgkinson, C. A., Pavan, W. J. & Arnheiter, H. (1997a) Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development*. [Online] 124 (12), 2377–2386.

Opdecamp, K., Nakayama, A., Nguyen, M. T., Hodgkinson, C. A., Pavan, W. J. & Arnheiter, H. (1997b) Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development (Cambridge, England)*. [Online] 124 (12), 2377–2386.

Van Otterloo, E., Li, W., Garnett, A., Cattell, M., Medeiros, D. M. & Cornell, R. A. (2012) Novel Tfap2-mediated control of soxE expression facilitated the

evolutionary emergence of the neural crest. *Development (Cambridge, England)*. 139 (4), 720–730.

Parichy, D. M., Mellgren, E. M., Rawls, J. F., Lopes, S. S., Kelsh, R. N. & Johnson, S. L. (2000) Mutational analysis of endothelin receptor b1 (rose) during neural crest and pigment pattern development in the zebrafish Danio rerio. *Developmental biology*. 227 (2), 294–306.

Parichy, D. M., Ransom, D. G., Paw, B., Zon, L. I. & Johnson, S. L. (2000) An orthologue of the kit-related gene fms is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, Danio rerio. *Development*. [Online] 127 (14), 3031–3044.

Parichy, D. M., Rawls, J. F., Pratt, S. J., Whitfield, T. T. & Johnson, S. L. (1999a) Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. *Development*. [Online] 126 (15), 3425–3436.

Parichy, D. M., Rawls, J. F., Pratt, S. J., Whitfield, T. T. & Johnson, S. L. (1999b) Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. *Development (Cambridge, England)*. [Online] 126 (15), 3425–3436.

Patton, E. E. & Zon, L. I. (2001) The art and design of genetic screens: zebrafish. *Nature Reviews Genetics*. [Online] 2 (12), 956–966.

Pearson, R. A., Dale, N., Llaudet, E. & Mobbs, P. (2005) ATP released via gap junction hemichannels from the pigment epithelium regulates neural retinal progenitor proliferation. *Neuron*. 46 (5), 731–744.

Pearson, R., Catsicas, M., Becker, D. & Mobbs, P. (2002) Purinergic and Muscarinic Modulation of the Cell Cycle and Calcium Signaling in the Chick Retinal Ventricular Zone. *J. Neurosci.* [Online] 22 (17), 7569–7579.

Peter, I. S., Faure, E. & Davidson, E. H. (2012) Predictive computation of genomic logic processing functions in embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (41), 16434–16442.

Pogenberg, V., Ogmundsdóttir, M. H., Bergsteinsdóttir, K., Schepsky, A., Phung, B., Deineko, V., Milewski, M., Steingrímsson, E. & Wilmanns, M. (2012) Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. *Genes & development*. 26 (23), 2647–2658.

Pollard, K., Dudoit, S. & Laan, M. van der (2004) Multiple Testing Procedures: R multtest Package and Applications to Genomics. U.C. Berkeley Division of Biostatistics Working Paper Series

Potterf, S. B., Furumura, M., Dunn, K. J., Arnheiter, H. & Pavan, W. J. (2000) Transcription factor hierarchy in Waardenburg syndrome: regulation of MITF expression by SOX10 and PAX3. *Human Genetics*. [Online] 107 (1), 1–6.

Ramos-Balderas, J. L., Carrillo-Rosas, S., Guzman, A., Navarro, R. E. & Maldonado, E. (2013) The zebrafish mutants for the V-ATPase subunits d, ac45, E, H and c and their variable pigment dilution phenotype. *BMC research notes*. 6 (1), 39.

Raposo, G. & Marks, M. S. (2007) Melanosomes--dark organelles enlighten endosomal membrane transport. *Nature reviews. Molecular cell biology*. 8 (10), 786–797.

Rawls, J. F. & Johnson, S. L. (2003) Temporal and molecular separation of the kit receptor tyrosine kinase's roles in zebrafish melanocyte migration and survival. *Developmental biology*. [Online] 262 (1), 152–161.

Reid, K., Turnley, A. M., Maxwell, G. D., Kurihara, Y., Kurihara, H., Bartlett, P. F. & Murphy, M. (1996) Multiple roles for endothelin in melanocyte development: regulation of progenitor number and stimulation of differentiation. *Development*. [Online] 122 (12), 3911–3919.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. & Smyth, G. K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 43 (7), e47 – .

Rodriguez-Cuenca, S., Barbarroja, N. & Vidal-Puig, A. (2015) Dihydroceramide desaturase 1, the gatekeeper of ceramide induced lipotoxicity. *Biochimica et biophysica acta*. 1851 (1), 40–50.

Rouzaud, F., Annereau, J.-P., Valencia, J. C., Costin, G.-E. & Hearing, V. J. (2003) Regulation of melanocortin 1 receptor expression at the mRNA and protein levels by its natural agonist and antagonist. *The FASEB journal*. [Online] 17 (14), 2154–2156.

Ruhrberg, C. & Schwarz, Q. (2010) In the beginning: Generating neural crest cell diversity. *Cell adhesion & migration*. 4 (4), 622–630.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. & Erlich, H. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239 (4839), 487–491.

Saito, H., Yasumoto, K.-I., Takeda, K., Takahashi, K., Fukuzaki, A., Orikasa, S. & Shibahara, S. (2002) Melanocyte-specific microphthalmia-associated transcription factor isoform activates its own gene promoter through physical

interaction with lymphoid-enhancing factor 1. *The Journal of biological chemistry*. 277 (32), 28787–28794.

Sandler, J. E. & Stathopoulos, A. (2016) Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. *Genetics*. 202 (4), 1575–1584.

Sato, S., Tanaka, M., Miura, H., Ikeo, K., Gojobori, T., Takeuchi, T. & Yamamoto, H. (2001) Functional conservation of the promoter regions of vertebrate tyrosinase genes. *The journal of investigative dermatology. Symposium proceedings / the Society for Investigative Dermatology, Inc. [and] European Society for Dermatological Research*. 6 (1), 10–18.

Schallreuter, K. U., Kothari, S., Chavan, B. & Spencer, J. D. (2008) Regulation of melanogenesis--controversies and new concepts. *Experimental dermatology*. 17 (5), 395–404.

Schallreuter, K., Wood, J., Pittelkow, M., Gutlich, M., Lemke, K., Rodl, W., Swanson, N., Hitzemann, K. & Ziegler, I. (1994) Regulation of melanin biosynthesis in the human epidermis by tetrahydrobiopterin. *Science*. 263 (5152), 1444–1446.

Schartl, M., Larue, L., Goda, M., Bosenberg, M. W., Hashimoto, H. & Kelsh, R. N. (2015) What is a vertebrate pigment cell? *Pigment cell & melanoma research*.

Sela-Donenfeld, D. & Kalcheim, C. (1999) Regulation of the onset of neural crest migration by coordinated activity of BMP4 and Noggin in the dorsal neural tube. *Development (Cambridge, England)*. [Online] 126 (21), 4749–4762.

Simon, J. D., Peles, D., Wakamatsu, K. & Ito, S. (2009) Current challenges in understanding melanogenesis: bridging chemistry, biological control, morphology, and function. *Pigment cell & melanoma research*. [Online] 22 (5), 563–579.

Singh, H., Khan, A. A. & Dinner, A. R. (2014) Gene regulatory networks in the immune system. *Trends in immunology*. 35 (5), 211–218.

Smith, J., Kraemer, E., Liu, H., Theodoris, C. & Davidson, E. (2008) A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Developmental biology*. 313 (2), 863–875.

Smolen, P., Baxter, D. A. & Byrne, J. H. (2000) Mathematical Modeling of Gene Networks. *Neuron*. 26 (3), 567–580.

Spiess, A.-N., Deutschmann, C., Burdukiewicz, M., Himmelreich, R., Klat, K., Schierack, P. & Rödiger, S. (2015) Impact of smoothing on parameter estimation in quantitative DNA amplification experiments. *Clinical chemistry*. 61 (2), 379–388.

Steingrímsson, E., Copeland, N. G. & Jenkins, N. A. (2004) Melanocytes and the microphthalmia transcription factor network. *Annual review of genetics*. 38365–411.

Sviderskaya, E. V, Easty, D. J. & Bennett, D. C. (1998) Impaired growth and differentiation of diploid but not immortal melanoblasts from endothelin receptor B mutant (piebald) mice. *Developmental dynamics : an official publication of the American Association of Anatomists*. 213 (4), 452–463.

Tabata, H., Kawamura, N., Sun-Wada, G.-H. & Wada, Y. (2008) Vacuolar-type H(+)-ATPase with the a3 isoform is the proton pump on premature melanosomes. *Cell and tissue research*. 332 (3), 447–460.

Tachibana, M., Perez-Jurado, L. A., Nakayama, A., Hodgkinson, C. A., Li, X., Schneider, M., Miki, T., Fex, J., Francke, U. & Arnheiter, H. (1994) Cloning of MITF, the human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1-p12.3. *Human molecular genetics*. [Online] 3 (4), 553–557.

Takesono, A., Moger, J., Farooq, S., Faroq, S., Cartwright, E., Dawid, I. B., Wilson, S. W. & Kudoh, T. (2012) Solute carrier family 3 member 2 (Slc3a2) controls yolk syncytial layer (YSL) formation by regulating microtubule networks in the zebrafish embryo. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (9), 3371–3376.

Theveneau, E. & Mayor, R. (2014) *Neural Crest Cells*. Elsevier.

Tong, Y. & Eigler, T. (2009) Transcriptional targets for pituitary tumor-transforming gene-1. *Journal of molecular endocrinology*. 43 (5), 179–185.

Tsujimura, T., Morii, E., Nozaki, M., Hashimoto, K., Moriyama, Y., Takebayashi, K., Kondo, T., Kanakura, Y. & Kitamura, Y. (1996) Involvement of transcription factor encoded by the mi locus in the expression of c-kit receptor tyrosine kinase in cultured mast cells of mice. *Blood*. [Online] 88 (4), 1225–1233.

Van der Vaart, M., van Soest, J. J., Spaink, H. P. & Meijer, A. H. (2013) Functional analysis of a zebrafish myd88 mutant identifies key transcriptional components of the innate immune system. *Disease models & mechanisms*. 6 (3), 841–854.

Waddington, C. H. (2014) *The Strategy of the Genes*. [Online]. Routledge.

Walker, W. P. & Gunn, T. M. (2010) Shades of meaning: the pigment-type switching system as a tool for discovery. *Pigment cell & melanoma research*. 23 (4), 485–495.

Wang, W.-D., Melville, D. B., Montero-Balaguer, M., Hatzopoulos, A. K. & Knapik, E. W. (2011) Tfap2a and Foxd3 regulate early steps in the development of the neural crest progenitor population. *Developmental biology*.

Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*. 10 (1), 57–63.

Watanabe, A., Takeda, K., Ploplis, B. & Tachibana, M. (1998) Epistatic relationship between Waardenburg syndrome genes MITF and PAX3. *Nature genetics*. 18 (3), 283–286.

Weger, B. D., Sahinbas, M., Otto, G. W., Mracek, P., Armant, O., Dolle, D., Lahiri, K., Vallone, D., Ettwiller, L., Geisler, R., Foulkes, N. S. & Dickmeis, T. (2011) The light responsive transcriptome of the zebrafish: function and regulation. *PloS one*. 6 (2), e17080.

Weiss, M. B., Abel, E. V, Dadpey, N. & Aplin, A. E. (2014) FOXD3 modulates migration through direct transcriptional repression of TWIST1 in melanoma. *Molecular cancer research : MCR*. 12 (9), 1314–1323.

Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J. L. & Meyerowitz, E. M. (2006) Genome-Wide Analysis of Gene Expression during Early Arabidopsis Flower Development. *PLoS Genetics*. 2 (7), e117.

Westerfield, M. (2000) *The zebrafish book*. [Online]

Wong, M. L. & Medrano, J. F. (2005) Real-time PCR for mRNA quantitation. *BioTechniques*. [Online] 39 (1), 75–85.

Wu, X. S., Rao, K., Zhang, H., Wang, F., Sellers, J. R., Matesic, L. E., Copeland, N. G., Jenkins, N. A. & Hammer, J. A. (2002) Identification of an organelle receptor for myosin-Va. *Nature Cell Biology*. [Online] 4 (4), 271–278.

Wu, X., Wang, F., Rao, K., Sellers, J. R. & Hammer, J. A. (2002) Rab27a Is an Essential Component of Melanosome Receptor for Myosin Va Thomas D Pollard (ed.). *Molecular Biology of the Cell*. [Online] 13 (5), 1735–1749.

Zhou, J. X., Aliyu, M. D. S., Aurell, E. & Huang, S. (2012) Quasi-potential landscape in complex multi-stable systems. *Journal of the Royal Society, Interface / the Royal Society*. 9 (77), 3539–3553.

Ziegler, I. (2003) The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. *Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society*. [Online] 16 (3), 172–182.

Appendices

APPENDIX 1

GRN mathematical model:

(Sox10) $$\frac{d}{dt}S = 2(g_S^{A|M,H} + g_S^{M|A,H}) - d_S S,$$

(Mitfa) $$\frac{d}{dt}M = 2(g_M^{S} + g_M^{Y|S}) - d_M M,$$

(Y) $$\frac{d}{dt}Y = 2g_Y^{M} - d_Y Y,$$

(Hdca1) $$\frac{d}{dt}H = 2g_H^{M} - d_H H,$$

(Tyrp1) $$\frac{d}{dt}T = 2g_T^{M} - d_T T,$$

(Dct) $$\frac{d}{dt}D = 2(g_D^{M|Z,S} + g_D^{Z|M,S}) - d_D D,$$

(Z [Sox9]) $$\frac{d}{dt}Z = 2g_Z^{B} - d_Z Z,$$

Where

$$A = \frac{A_0}{1+e^{\beta\ (t_A - t)}}, \quad B = \frac{B_0}{1+e^{\beta\ (t-t_2)}} - \frac{B_0}{1+e^{\beta\ (t-t_1)}},$$

$$r_{AH} = \frac{\alpha_0\phi_1 A}{\alpha_1\phi_1 + \alpha_0\phi_1 A + \alpha_1\phi_0 H}, \quad r_{MH} = \frac{\beta_{\ 0}\xi_1 M}{\beta_{\ 1}\xi_1 + \beta_{\ 0}\xi_1 M + \beta_{\ 1}\xi_0 H},$$

$$g_S^{A|M,H} = g_S r_{AH}(1 - r_{MH}/2), \quad g_S^{M|A,H} = g_S r_{MH}(1 - r_{AH}/2),$$

$$g_M^{S} = g_M \frac{(\gamma_0^{(1)}\gamma_1^{(2)} + \gamma_0^{(2)}\gamma_1^{(1)}\frac{k_1}{k_2}S)S}{(\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_0^{(1)}\gamma_1^{(2)}S + \gamma_0^{(2)}\gamma_1^{(1)}\frac{k_1}{k_2}S^2)}, \quad g_M^{Y|S} = g_M \frac{\gamma_1^{(1)}\gamma_1^{(2)}\frac{\sigma_0 He(Y-Y^*)Y}{\sigma_1+\sigma_0 He(Y-Y^*)Y}}{(\gamma_1^{(1)}\gamma_1^{(2)} + \gamma_0^{(1)}\gamma_1^{(2)}S + \gamma_0^{(2)}\gamma_1^{(1)}\frac{k_1}{k_2}S^2)}$$

$$g_Y^{M} = g_Y \frac{\delta_0 He(M - M^*)M}{\delta_1 + \delta_0 He(M - M^*)M},$$

$$g_H^{M} = g_H \frac{\theta_0 M}{\theta_1 + \theta_0 M},$$

$$g_T^{M} = g_T \frac{\mu_0 M}{\mu_1 + \mu_0 M},$$

$$r_M = \frac{\lambda_0 M}{\lambda_1 + \lambda_0 M}, \quad r_Z = \frac{\varepsilon_0 Z}{\varepsilon_1 + \varepsilon_0 Z}, \quad r_S = \frac{\nu_0 S}{\nu_1 + \nu_0 S},$$

$$g_D^{M|Z,S} = g_D r_M(1 - r_Z/2)(1 - r_S), \quad g_D^{Z|M,S} = g_D r_Z(1 - r_M/2)(1 - r_S),$$

$$g_Z^{B} = g_Z \frac{\rho_0 B}{\rho_1 + \rho_0 B}.$$

Here *He*(·) is the Heaviside function. The initial data at time t_0 = 0 and known parameters are

$$
\begin{array}{lllllll}
Z(t_0)=0, & t_1=12, & \beta_1=1.2, & \gamma_0^{(2)}=1, & \delta_1=0.5, & \varepsilon_0=0.2, & d_Y=0.2,\\
S(t_0)=0, & t_2=24, & \phi_0=2, & \gamma_1^{(1)}=1, & \theta_0=0.1, & \varepsilon_1=1.3, & g_H=0.3,\\
M(t_0)=0, & A_0=1, & \phi_1=0.1, & \gamma_1^{(2)}=0.8, & \theta_1=0.5, & g_Z=0.1, & d_H=0.03,\\
Y(t_0)=0, & B_0=1, & \xi_0=1.6, & k_1=1, & \mu_0=0.1, & d_Z=0.2, & g_T=3.1,\\
H(t_0)=0, & \rho_0=1, & \xi_1=1.1, & k_2=1, & \mu_1=1.3, & g_S=0.3, & d_T=0.2,\\
T(t_0)=0, & \rho_1=0.5, & Y^*=0.01, & \sigma_0=1.6, & \lambda_0=1, & d_S=0.3, & g_D=0.3,\\
D(t_0)=0, & \alpha_0=2.2, & M^*=0.01, & \sigma_1=1.1, & \lambda_1=1, & g_M=0.3, & d_D=0.1,\\
\beta=2, & \alpha_1=1, & \gamma_0^{(1)}=1.5, & \delta_0=1.6, & \nu_0=1, & d_M=0.3, & \\
t_A=12, & \beta_0=1.3, & & & \nu_1=1.2, & g_Y=0.15, &
\end{array}
$$

all taken from Greenhill *et al.* (2011).

APPENDIX 2

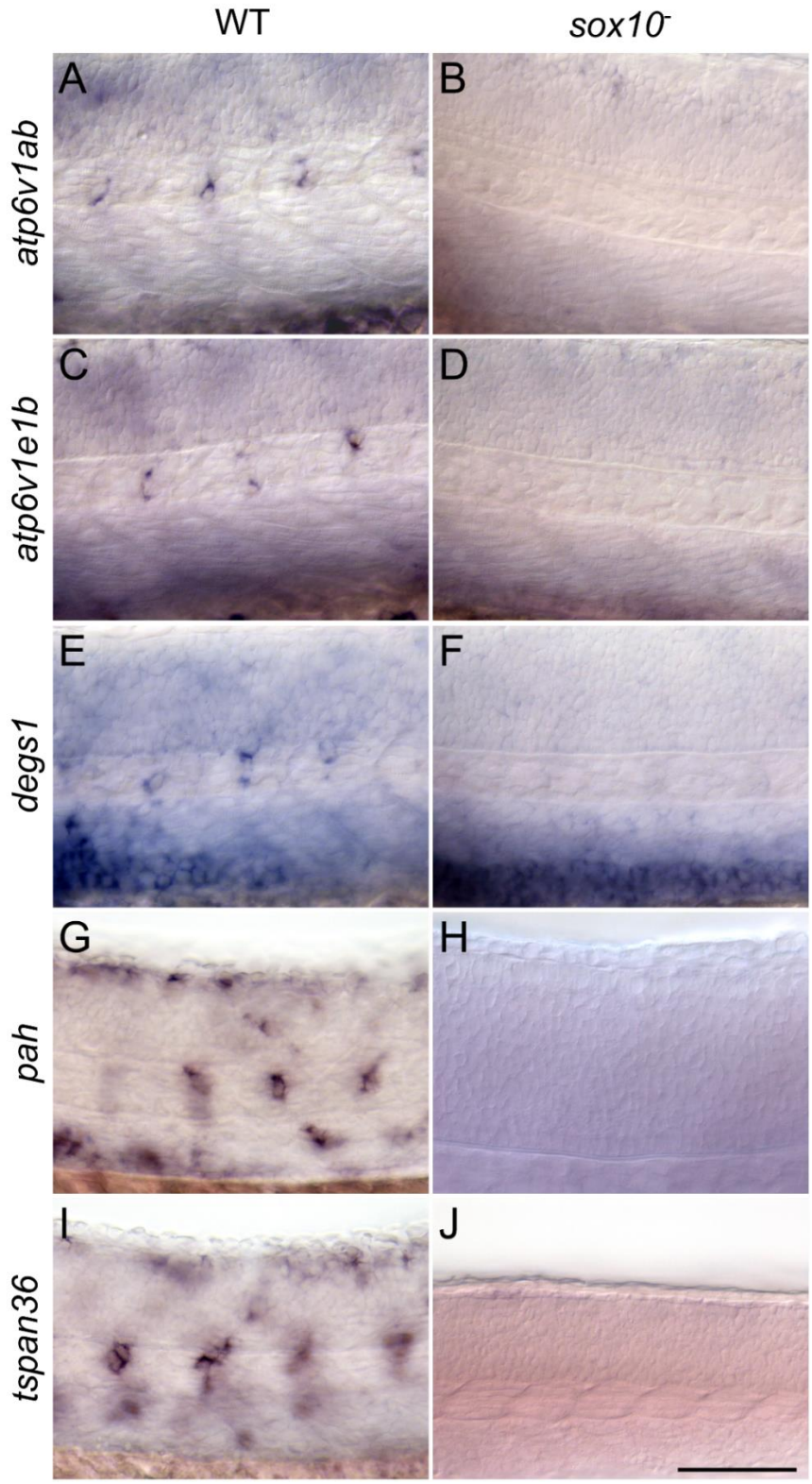

Figure 40. ISH expression pattern of the 5 verified melanocyte genes in 30 hpf wild type and *sox10* mutant embryos. All images are lateral views of the mid-trunk embryo. All 5 genes (*atp6v1ab*, *atp6v1e1b*, *degs1*, *pah* and *tspan36*) are expressed in melanocytes of wild type embryos (A, C, E, G and I)) and absent in *sox10* mutant embryos (B, D, F, H and J). Scale: 100 µm

APPENDIX 3

R script used to analyse the combined microarray dataset:

```
#*****************************************
#
#________ Microarrays analysis script_______
#____________Alberto Lapedriza_____________
#
#*****************************************

# Clear workspace
rm(list=ls())

# Close any open graphics devices
graphics.off()

#Loading libraries
library(affy)
library(AnnotationDbi)
library(affyPLM)
library(genefilter)
library(limma)
library(e1071)
library(MASS)
library(multtest)
library(annaffy)
library(zebrafish.db)
library(GO.db)
library(gplots)

#__________LOADING THE DATA__________

#Create a special DataFrame which content the experiment
description file
total.expt<-
read.AnnotatedDataFrame("bothSets.txt",header=TRUE,row.names=1
,sep="\t")

#Load all the .CELL files into an AffyBatch
total.data<-
ReadAffy(filenames=rownames(pData(total.expt)),phenoData=total
.expt,verbose=TRUE)

#_________PRE-PROCESSING__________

#Normalise the data using the RMA.2 background correction
method;quantile normalisation;no perfect match evaluation and
a median polish summarisation.
prepro<-
threestep(total.data,background.method="RMA.2",normalize.metho
d="quantile",summary.method="median.polish")

#___________FILTERING____________
#Filter genes that are absent in all samples
```

```
#Performs the Wilcoxon signed rank-based gene expression
presence/absence detection algorithm
eset.mas5calls<-mas5calls(total.data)

#Create a funcion
mas5callsfilter<-function(cutoff="A",number){
      function(x){
            sum(x==cutoff) !=number
      }
}

#Filtering the absent genes of the data
mas5callsfunction<-
mas5callsfilter(number=dim(eset.mas5calls)[2])
m5cfun<-filterfun(mas5callsfunction)

m5csub<-genefilter(eset.mas5calls,m5cfun)
sum(m5csub)

#Subset of the samples after apply the absent genes filter
filtered<-exprs(prepro[m5csub,])
dim(filtered) #11732

#___________STATISTICAL TESTING____________

#Create the design matrix
design<- model.matrix(~ 0+factor(c(1,1,1,1,1,2,2,2,2,2)))
colnames(design)<- c("Mutant","WT")

#Fit a linar model to the data set
fit<-lmFit(filtered,design)

#Generate a contrast matrix, because we want to identify genes
that are differentially expressed between the Mutant group and
the WT
contrast.matrix<-makeContrasts(Mutant-WT, levels=design)

#Fit the contrast to the linear model and compute the t-
statistics using function eBayes
fit2<-contrasts.fit(fit,contrast.matrix)
fit2<-eBayes(fit2)

#Correct the raw p-values in order to control the false positive
rate. Using procedure Benjamini and Hochberg
multadjust<-mt.rawp2adjp(fit2$p.value, proc=c("BH"))

#Restore original gene order
eBpvalues<-multadjust$adjp[order(multadjust$index),]
rownames(eBpvalues)<-rownames(fit2$p.value)
dim(eBpvalues) #11732

#___________RESULTS GENE LIST____________

#Extract all the differenctially expressed gene with a p-value
smaller than 0.05
```

```
#USING THE MULTITEST CORRECTED P-VALUE
results<-NULL
j<-1
for(i in 1:dim(eBpvalues)[1]){
      if(eBpvalues[i,2]<=0.05){
            results<-rbind(results,eBpvalues[i,])
            row.names(results)[j]<-row.names(eBpvalues)[i]
            j<-j+1
      }
}

if(is.null(dim(results))){
      cat("\nTHERE'S NO DIFFERENCTIALLY EXPRESSED GENES\n")
      cat("\n-----END OF THE SCRIPT-----\n")
      stop()
}

#Save the result in a csv file
write.table(results,file="results.csv",sep=",",col.names=NA,ro
w.names=TRUE)
dim(results)#113

#___________ANNOTATION___________

symbols <- rep(NA, dim(results)[1])

for(i in 1:dim(results)[1]){
      symbols[i]<-
getText(aafSymbol(rownames(results)[i],"zebrafish.db"))

      if(symbols[i]==""){
            symbols[i]<-rownames(results)[i]
      }
}

#Assign to the gene list row names the names of the genes
instead of the microarray ID
#genelist has now the gene name as row names and then 3 colums
with the raw Pvalue, the corrected Pvalue and the t statistics
rownames(results) = make.names(symbols, unique=TRUE)
results=results[order(results[,2],decreasing=FALSE),]

write.table(results,file="annotated_genelist.csv",sep="\t",col
.names=NA)
```

APPENDIX 4

Perl script used to locate the Sox10 binding sites. The one to locate the Mitfa binding sites is similar but with the Mitfa binding motifs instead of the Sox10 ones:

```
#!/usr/bin/perl -w

use strict;
use Bio::SeqIO;

# get command-line arguments, or die with a usage statement
my $usage = "testScript.pl infile outfile\n";
my $infile = shift or die $usage;

# create a SeqIO object to read in the file
my $seq_in = Bio::SeqIO-> new(-file => "<$infile");
#my $seq_out = Bio::SeqIO-> new(-file => ">$outfile");

my $motifS1="CTCAAAG";
my $motifS3="GATTGTA";

# write each entry in the input file to the output file
#First instance promotors with just the S1 motif
my $count=0;
my @seq_array;
while (my $seq=$seq_in->next_seq){

     if(($seq->seq   =~   /$motifS1/)   &&   ($seq->seq   =~
/$motifS3/)){
          $count++;
          print $count,"\n";
          push(@seq_array,$seq->id);
     }
     else{
     }
}

my $id = join("\n",@seq_array);

open FILE, ">idsWithS1andS3.txt" or die $!;
print FILE $id;
close FILE;

exit;
```

APPENDIX 5

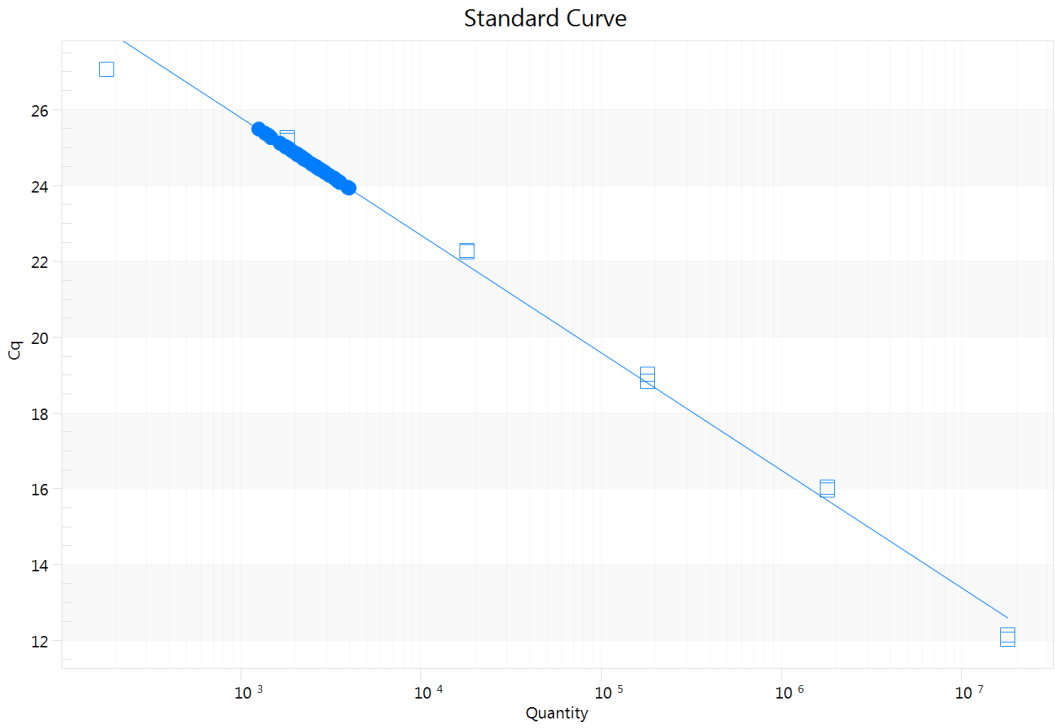

Figure 41. *sox10* standard curve. Equation y=-3.1x+35.1 R^2=0.992 Efficiency=110.11%

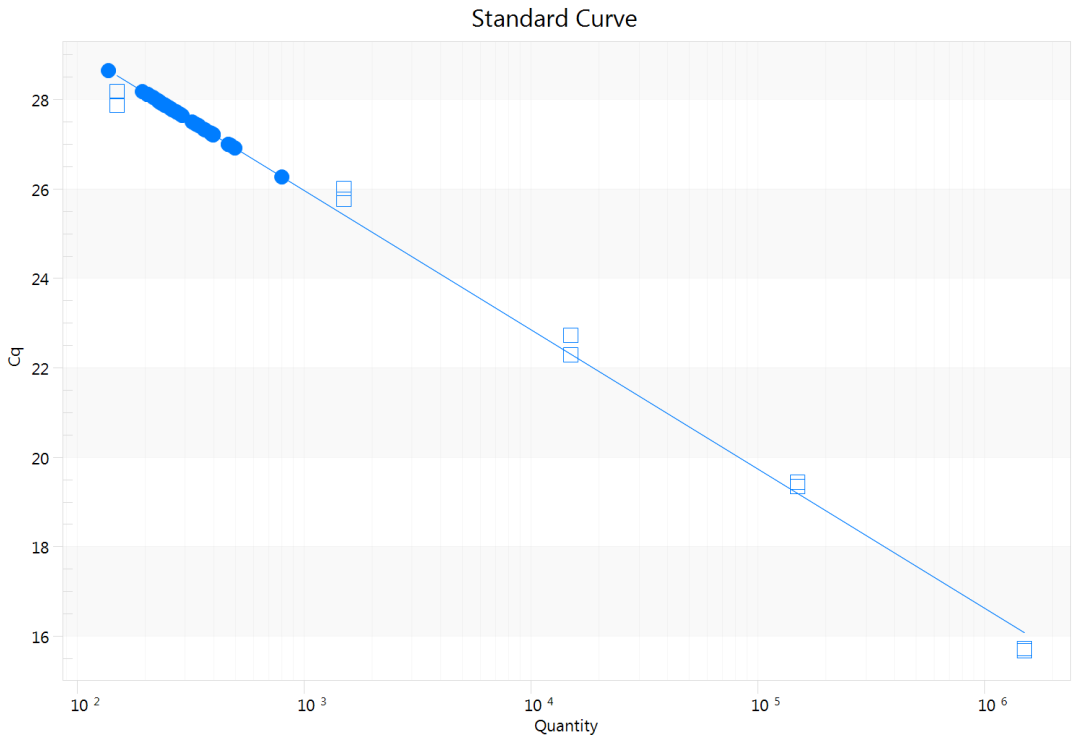

Figure 42. *kit* standard curve. Equation: y=-3.12x+35.32, R^2=0.992, Efficiency=109.39%

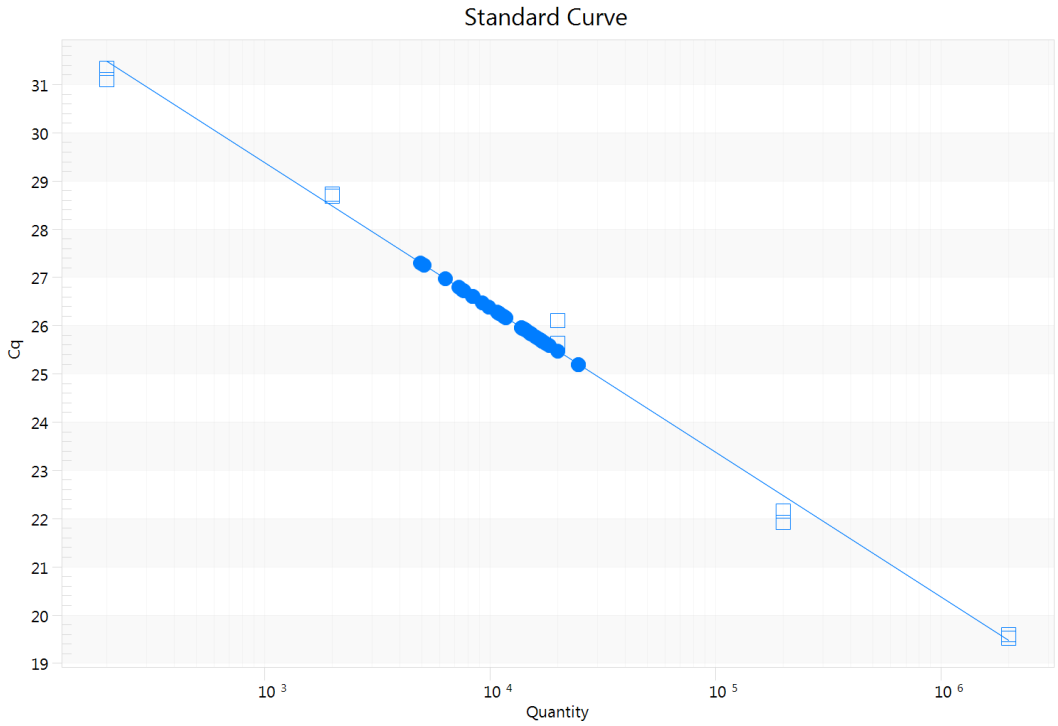

Figure 43. *mitfa* standard curve. Equation: y=-3x+38.38, R^2=0.994, Efficiency=115.39%

APPENDIX 6

Data smoothing algorithm:

```
%Import the copy numbers data from Excel giving the file name
[mitfacopies,mitfaStDev]=importGenerali('mitfadata');
[sox10copies,sox10StDev]=importGenerali('sox10data');
[kitcopies,kitStDev]=importGenerali('kitdata');
[dctcopies,dctStDev]=importGenerali('dctdata');

%Import the cell number data from Excell
[Stage,cellnumb,cellnumbStDev]=importGeneralii('cellcount');

%choose the smoothing method
method='sgolay';

%smooth the copies data
mitfacopiessmooth=smooth(mitfacopies,method);
sox10copiessmooth=smooth(sox10copies,method);
kitcopiessmooth=smooth(kitcopies,method);
dctcopiessmooth=smooth(dctcopies,method);

%transform all the negative values in the copie number data matrix
to 0
mitfacopiessmooth(mitfacopiessmooth<0)=0;
sox10copiessmooth(sox10copiessmooth<0)=0;
```

```
kitcopiessmooth(kitcopiessmooth<0)=0;
dctcopiessmooth(dctcopiessmooth<0)=0;

%smooth the cell count data
cellnumbsmooth=smooth(cellnumb,method);
%round each element of cellnumbsmooth to the nearest integer
cellnumbsmooth = round(cellnumbsmooth);


% write a csv file with the new smooth data
%matrix =
horzcat(stages,sox10cellsmooth,kitcellsmooth,mitfadatasmooth,dctcell
smooth);
%csvwrite('smoothcopiesdensity',matrix);

% plot the copy number data
figure;
subplot(2,2,1)
plot(Stage,sox10copies,':');
hold on
errorbar(Stage,sox10copiessmooth,sox10StDev,'-x');
title('sox10');
xlabel('stages (hpf)');
ylabel('copy number');
xlim([15 100])
ylim([0 inf])
hold off;

subplot(2,2,2)
plot(Stage,kitcopies,':');
hold on
errorbar(Stage,kitcopiessmooth,kitStDev,'-x');
title ('kit');
xlabel('stages (hpf)');
ylabel('copy number');
xlim([15 100])
ylim([0 inf])
hold off;

subplot(2,2,3)
plot(Stage,mitfacopies,':');
hold on
errorbar(Stage,mitfacopiessmooth,mitfaStDev,'-x');
title ('mitfa');
xlabel('stages (hpf)');
ylabel('copy number');
xlim([15 100])
ylim([0 inf])
hold off;

subplot(2,2,4)
plot(Stage,dctcopies,':');
hold on
errorbar(Stage,dctcopiessmooth,dctStDev,'-x');
title ('dct');
xlabel('stages (hpf)');
ylabel('copy number');
xlim([15 100])
ylim([0 inf])
hold off;
```

```
%legend
legend('Original data', ['Smoothed (',method,')
data'],'Location','SouthEast');

%save the figure
print('smoothCopyNumbGraphs','-dtiff','-r400');

% plot the data of cell number
figure;
plot(Stage,cellnumb,':');
hold on
errorbar(Stage,cellnumbsmooth,cellnumbStDev,'-x');
title ('cell number');
xlabel('stages (hpf)');
ylabel('cell number');
xlim([15 100])
hold off;

%legend
legend('Original data', ['Smoothed (',method,')
data'],'Location','SouthEast');

%save the figure
print('smoothCellNumbGraphs','-dtiff','-r400');

%divide the smoothed copy number data by the smoothed cell number
data to
%get copies per cell for mitfa and dct as are the only ones I'm
gonna use
mitfacopiespercell = (mitfacopiessmooth./cellnumbsmooth);
dctcopiespercell = (dctcopiessmooth./cellnumbsmooth);

%transform all the negative values in the data matrix to 0
mitfacopiespercell(mitfacopiespercell<0)=0;
dctcopiespercell(dctcopiespercell<0)=0;

%The St. Dev. of the copy numbers has to be divided by the St.Dev of
the
%cell number. To do that we applied the following formula:
mitfacopiespercellStDev=((cellnumbsmooth.*mitfaStDev)+(mitfacopiessm
ooth.*cellnumbStDev))./(cellnumbsmooth.^2);
dctcopiespercellStDev=((cellnumbsmooth.*dctStDev)+(dctcopiessmooth.*
cellnumbStDev))./(cellnumbsmooth.^2);

mitfacopiespercellStDev(isnan(mitfacopiespercellStDev))=0;
dctcopiespercellStDev(isnan(dctcopiespercellStDev))=0;


%Plot the copy number per cell data of mitfa and dct as these are
the only
%ones I'm going to use
figure;

subplot(2,1,1)
errorbar(Stage,mitfacopiespercell,mitfacopiespercellStDev,'-x');
title('mitfa');
xlabel('stages (hpf)');
ylabel('copies per cell');
```

```
xlim([15 100])
ylim([0 inf])

subplot(2,1,2)
errorbar(Stage,dctcopiespercell,dctcopiespercellStDev,'-x');
title('dct');
xlabel('stages (hpf)');
ylabel('copies per cell');
xlim([15 100])
ylim([0 inf])

%save the figure
print('copiespercell','-dtiff','-r400');

%write the stages, copies per cell and its St Dev to a csv file for
later
%use for mitfa and dct
csvwrite('mitfacopiespercell.csv',(horzcat(Stage,mitfacopiespercell,
mitfacopiespercellStDev)));
csvwrite('dctcopiespercell.csv',(horzcat(Stage,dctcopiespercell,dctc
opiespercellStDev)));
```

APPENDIX 7

Mathematical model equations:

```
function dy = RHS(t,y)
dy = zeros(size(y));
RHSaux;
dy(2)=2*(gSAMH+gSMAH)-y(49)*y(2);
dy(3)=2*(gMS+gMYS)-y(51)*y(3);%mitfa
dy(4)=2*gYM-y(53)*y(4);
dy(5)=2*gHM-y(55)*y(5);
dy(6)=2*gTM-y(57)*y(6);
dy(7)=2*(gDMZS+gDZMS)-y(59)*y(7);
dy(1)=2*gZB-y(47)*y(1);
```

Auxiliary functions:

```
HY=(1)./(1+exp(50*(y(24)-y(4))));
HM=(1)./(1+exp(50*(y(25)-y(3))));
A=(y(12))./(1+exp(y(8)*(y(9)-t)));
B=-(y(13))./(1+exp(y(8)*(y(10)+y(11)-
t)))+(y(13))./(1+exp(y(8)*(y(10)-t)));
rAH=(y(16)*y(21)*A)./(y(17)*y(21)+y(16)*y(21)*A+y(17)*y(20)*y(5));
rMH=(y(18)*y(23)*y(3))./(y(19)*y(23)+y(18)*y(23)*y(3)+y(19)*y(22)*y(
5));
gSAMH=y(48)*rAH*(1-rMH/2);
gSMAH=y(48)*rMH*(1-rAH/2);
```

```
gMS=y(50)*((y(26)*y(29)+y(27)*y(28)*(y(30))./(y(31))*y(2))*y(2))./((
y(28)*y(29)+y(26)*y(29)*y(2)+y(27)*y(28)*(y(30))./(y(31))*y(2)^2)) ;
gMYS=y(50)*(y(28)*y(29)*(y(32)*HY*y(4))./(y(33)+y(32)*HY*y(4)))./((y
(28)*y(29)+y(26)*y(29)*y(2)+y(27)*y(28)*(y(30))./(y(31))*y(2)^2)) ;
gYM=y(52)*(y(34)*HM*y(3))./(y(35)+y(34)*HM*y(3));
gHM=y(54)*(y(36)*y(3))./(y(37)+y(36)*y(3));
gTM=y(56)*(y(38)*y(3))./(y(39)+y(38)*y(3));
rM=(y(40)*y(3))./(y(41)+y(40)*y(3));
rZ=(y(44)*y(1))./(y(45)+y(44)*y(1));
rS=(y(42)*y(2))./(y(43)+y(42)*y(2));
gDMZS=y(58)*rM*(1-rZ/2)*(1-rS);
gDZMS=y(58)*rZ*(1-rM/2)*(1-rS);
gZB=y(46)*(y(14)*B)./(y(15)+y(14)*B);
```

APPENDIX 8

Parameter optimisation algorithm:

```
global selec y options tspan Ymeas YmeasStD YmeasRelStD;

fillm=1;

%Import the copy number per cell data from the mitfa and dct csv
file produced in the smoothing process
[stages,mitfacopies,mitfacopiesStDev]=importGeneraliii('mitfacopiesp
ercell');
[stages,dctcopies,dctcopiesStDev]=importGeneraliii('dctcopiespercell
');


tspan=stages;

tspan=[20:1:30, 32:2:52, 56:4:96]';
dctcopies = interp1(stages,dctcopies,tspan);
dctcopiesStDev = interp1(stages,dctcopiesStDev,tspan);
mitfacopies = interp1(stages,mitfacopies,tspan);
mitfacopiesStDev = interp1(stages,mitfacopiesStDev,tspan);

Ymeas=(2/3000)*vertcat(dctcopies,mitfacopies);
YmeasStD=(2/3000)*vertcat(dctcopiesStDev/3,mitfacopiesStDev/6)/12;
YmeasRelStD=YmeasStD./Ymeas;

selec=[9:59];

% load initial parameters and
% plot trajectories for this IC
odeSolver;

%look for the best parameters using the Greenhill parameter values
as first
%guess
options = odeset('RelTol',1e-2); % sets the tolerance
```

```
lsoptions = optimset('Display','iter');

% load y from file
load parameterfilemm;

y(8)=50;  % sharp A,B rectangles

para0=y(selec);

para0=para0.*exp(random('Normal',0,0.00001,1,length(selec)));


lb=.00*ones(size(para0));
ub=1e2*ones(size(para0));

[para,resnorm,residual,exitflag] = lsqnonlin(@PEerror,
para0,lb,ub,lsoptions);
para0=y;

%para0(selec)=exp(para);
para0(selec)=para;

paraodeSolver;
save parameterfile y

hold on;

ti=tspan;tl=length(ti);
plot(ti,((3000/2)*Ymeas(1:tl)),'b.');
plot(ti,((3000/2)*Ymeas(tl+1:end)),'g.');

[stages,mitfacopies,mitfacopiesStDev]=importGeneraliii('mitfacopiesp
ercell');
[stages,dctcopies,dctcopiesStDev]=importGeneraliii('dctcopiespercell
');
Ymeas=(2/3000)*vertcat(dctcopies,mitfacopies);
YmeasStD=(2/3000)*vertcat(dctcopiesStDev,mitfacopiesStDev);

ti=stages;tl=length(ti);
%plot(ti,Ymeas(1:tl),'bo');
errorbar(ti,((3000/2)*Ymeas(1:tl)),((3000/2)*YmeasStD(1:tl)),'xb');
%plot(ti,Ymeas(tl+1:end),'go');
errorbar(ti,((3000/2)*Ymeas(tl+1:end)),((3000/2)*YmeasStD(tl+1:end))
,'xg');
ylim([0 inf])
hold off;

print('optimised_parametersHartmutforThesis','-dpng');%,'-r400');
```

APPENDIX 9

New parameter set after the optimisation algorithm:

```
y=[       0                                    0.9034
          0                                    0.5804
          0                                    2.4771
          0                                    1.2058
          0                                    1.7523
          0                                    0.0290
          0                                    1.2133
    50.0000                                    0.0019
    17.3836                                    1.2893
    19.7444                                    0.0337
    21.5732                                   51.8319
     0.7879                                   15.3247
     1.5166                                    1.5921
     1.5163                                    0.1899
     0.3125                                    1.6547
     1.7649                                    0.1873
     1.5014                                    0.0443
     0.5054                                    0.6880
     5.5983                                    3.2927
     7.5692                                   22.8155
     0.0537                                    3.0681
     0.7246                                    0.0204
     2.8502                                    0.0998
     0.1452                                    0.3465
     0.5032                                    0.0000
     2.5221                                    0.0165
     3.1304                                    0.3770
     2.5167                                    1.4234
     0.8393                                    0.0736
     3.1280                                         ];
```